GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:22:10 ; Search time 19.89 Seconds
(without alignments)
528.810 Million cell updates/sec

Title: US-09-382-242-33
Perfect score: 942
Sequence: 1 MSLNKHSWMDMIIFILSFSF.....AVFSFLPFYIRDYFKSYTKR 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS6/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	942	100.0	184	18	W23069	Staphylothermus ma
2	90.5	9.6	442	19	W98870	H. pylori GHPO 171
3	90.5	9.6	442	19	W71557	Helicobacter polyp
4	89.5	9.5	258	21	G14660	Arabidopsis thalia
5	89.5	9.5	258	21	G14825	Arabidopsis thalia
6	89.5	9.5	342	21	G14659	Arabidopsis thalia
7	89.5	9.5	342	21	G14824	Arabidopsis thalia
8	89.5	9.5	356	21	G14823	Arabidopsis thalia
9	89.5	9.5	485	20	Y35687	Chlamydia pneumoni
10	87	9.2	249	19	W79385	Staphylococcus aur
11	87	9.2	412	21	Y70156	Staphylococcus aur
12	87	9.2	435	19	W98458	H. pylori GHPO 710
13	86.5	9.2	214	21	G14661	Arabidopsis thalia
14	86.5	9.2	397	21	G36489	Arabidopsis thalia
15	86	9.1	428	20	Y35302	Amino acid sequenc
16	86	9.1	432	18	W14081	S.thermophilus exo
17	86	9.1	473	18	W22180	S.thermophilus exo
18	85.5	9.1	212	21	G50404	Arabidopsis thalia
19	85.5	9.1	240	21	G50403	Arabidopsis thalia
20	85.5	9.1	266	21	G50402	Arabidopsis thalia
21	85.5	9.1	358	20	Y36803	Chlamydia trachoma
22	85	9.0	279	21	Y88627	H. influenzae dime
23	84.5	9.0	212	21	G11179	Arabidopsis thalia
24	84.5	9.0	240	21	G11178	Arabidopsis thalia
25	84.5	9.0	266	21	G11177	Arabidopsis thalia
26	84.5	9.0	284	21	G36490	Arabidopsis thalia
27	84.5	9.0	397	16	W13742	Urea transporter p
28	84.5	9.0	430	20	W88344	Salmonella enteric
29	84	8.9	363	19	W75788	Staphylococcus aur
30	83.5	8.9	406	21	Y81562	Streptococcus pneu
31	83.5	8.9	649	21	B12140	Hydrophobic domain
32	83	8.8	335	20	W92951	WO9905287 Seq ID 1
33	83	8.8	360	16	R80953	Recombinant high a
34	82.5	8.8	378	21	B53159	Macaca mulatta rha
35	82.5	8.8	474	19	W58862	T. halophilus xyul
36	82	8.7	840	21	B11037	S. aureus mprF pro
37	81.5	8.7	355	21	Y75278	Neisseria meningit
38	81.5	8.7	487	12	R14149	Human alpha 2 beta
39	81.5	8.7	487	18	W11804	Human alpha-2b adr
40	81.5	8.7	1137	22	B36936	Integrin alpha cha
41	81.5	8.7	1141	20	Y41752	Human PRO768 prote
42	81.5	8.7	1141	21	B44308	Human PRO768 (UNQ4
43	81	8.6	395	21	Y81541	Streptococcus pneu
44	80	8.5	504	21	Y81574	Streptococcus pneu
45	79	8.4	421	18	W20180	H. pylori putative

ALIGNMENTS

RESULT 1
W23069
ID W23069 standard; Protein; 184 AA.
XX
AC W23069;
XX
DT 17-FEB-1998 (first entry)
XX
DE Staphylothermus marinus esterase F1-12LC.
XX
KW Esterase; thermostable enzyme; ester; chiral compound; cheese;
KW pulp; paper; lignin removal; sugar; lignocellulose;
KW disease resistance; feedstuff.
XX
OS Staphylothermus marinus strain F1.
XX
FH Key Location/Qualifiers
FT Misc-difference 176
FT /note= "encoded by CAG"
XX
PN WO9730160-A1.
XX
PD 21-AUG-1997.
XX
PF 11-FEB-1997; 97WO-US02039.
XX
PR 16-FEB-1996; 96US-0602359.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Callen W, Kosmotka A, Link S, Maffia AM, Murphy D;
PI Reid J, Robertson DE, Swanson RV, Warren PV;
XX
DR WPI; 1997-425035/39.

DR N-PSDB; T79321.
XX
PT Nucleic acid encoding heat stable esterase from thermophilic
PT bacteria - which is active in organic solvents, useful in cheese or
PT paper manufacture, and to study plant resistance to disease
XX
PS Claim 19; Page 53-54; 113pp; English.
XX
CC This protein comprises the thermostable esterase F1-12LC of
CC Staphylothermus marinus F1, a thermophilic sulphur archaea that
CC grows optimally at 85 deg C and pH 6.5. Newly identified
CC polynucleotides (T79321-30) encoding claimed esterases (W23069-77,
CC W23088) were recovered from genomic gene libraries. They can be
CC used for recombinant production of the enzymes in host cells, and as
CC probes to identify related sequences. The esterases are stable at
CC high temperature and in organic solvents, making them superior for
CC use in production of pure chiral compounds used in pharmaceutical,
CC agricultural and other chemical industries. A method is claimed
CC for transferring an amino group from an amino acid to an alpha-keto
CC acid using a claimed esterase. The enzymes may also be useful as
CC ripening starters in cheese making, in lignin removal in paper and
CC pulp manufacture, in carbohydrate derivative synthesis, in
CC fermentable sugar production from lignocellulosic waste, in the
CC study of plant wall structure, plant resistance to disease and
CC organic matter decomposition and to select plants bred for
CC production of highly degradable animal feeds.
XX
SQ Sequence 184 AA;

Query Match 100.0%; Score 942; DB 18; Length 184;
Best Local Similarity 100.0%; Pred. No. 4.4e-102;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MSLNKHSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mslnkhswmdmiifilsfsfpltmialaismsswfniwnnalsdlghavkssvapifnlg 60

Qy     61 LAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSII 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 laiggilivivglrnlyswsrvkgsliismgvflnligvfdevygwihflvsvlfflsii 120

Qy    121 AYFIAISILDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYIRDYFKS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ayfiaisildkswiavlliighiamwylhfaseiprgaaipellavfsflpfyirdyfks 180

Qy    181 YTKR 184
          ||||
Db    181 ytkr 184
```

RESULT 2
W98870
ID W98870 standard; Protein; 442 AA.
XX
AC W98870;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1719 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; X14589.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 8; Page 1972-1974; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 442 AA;

Query Match 9.6%; Score 90.5; DB 19; Length 442;
Best Local Similarity 22.1%; Pred. No. 0.027;
Matches 45; Conservative 46; Mismatches 54; Indels 59; Gaps 11;

```
Qy     18 FSFPLTMIALAISMSSWFNIWNNALSD-LGHAVKSSVAPIFNLGLAIGGILIVIVGLRNL 76
          |:: |  :  ::|:    ||   |::|   : :||::       : : |||| :|    :
Db    211 ftyslgqvffslsiglginityaavtdktqnllksti------wvvlsgilislvaglmi 264

Qy     77 YSWSRVKGSLIISMG---VFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFI--AISILDK 131
          :::    |:  :| |   :| :|  || :: | |  |||:|| |::    |   :::|:
Db    265 ftfvfeyga-nvsqgtgliftslpvvfgqm-gaigilvsilfllalafagitstvallep 322

Qy    132 S----------------W--IAVLLIIGHIAMWYLH------------------FASE- 153
          |                |  :|:: ::| : :: |:|                   |||
Db    323 svmylteryqysrfkvtwglvalifvvgvvlifslhkdykdyltffekslfdwldfasst 382

Qy    154 --IPRGAAIPELLAVFSFLPFYIR 175
            :| |      :| | |: : ::
Db    383 iimplgg-----matfifmgwvlk 401
```

RESULT 3
W71557
ID W71557 standard; Protein; 442 AA.
XX
AC W71557;
XX
DT 09-NOV-1998 (first entry)
XX
DE Helicobacter polypeptide GHPO 1719.
XX
KW GHPO 1719; infection; therapy; diagnosis; vaccine; gastritis;
KW ulcer.
XX
OS Helicobacter pylori.
XX
PN WO9821225-A1.
XX
PD 22-MAY-1998.
XX
PF 14-NOV-1997; 97WO-US21353.
XX
PR 29-JUL-1997; 97US-0902615.
PR 14-NOV-1996; 96US-0749051.
PR 01-APR-1997; 97US-0831309.

PR 01-APR-1997; 97US-0833457.
PR 01-APR-1997; 97US-0834705.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C;
PI Odenbreit S, Tomb J;
XX
DR WPI; 1998-297855/26.
DR N-PSDB; V52092.
XX
PT Helicobacter polynucleotide and polypeptide sequences - useful to
PT treat or prevent gastrointestinal infection
XX
PS Claim 1; Page 339-340; 362pp; English.
XX
CC This claimed Helicobacter pylori polypeptide, designated GHPO 1719,
CC can be used in vaccination methods for preventing or treating
CC Helicobacter infection. 85 Helicobacter polypeptides (see
CC W71474-W71558) are claimed, as well as isolated polynucleotides
CC (see V52009-93) that encode them. The invention also provides:
CC methods for producing these Helicobacter polypeptides in
CC recombinant host systems, and related expression cassettes, vectors
CC and transformed or transfected host cells; live vaccine vectors
CC that contain the polynucleotides of the invention and which can be
CC used to prevent or treat Helicobacter infection; therapeutic and/or
CC prophylactic methods involving administration of polynucleotide
CC molecules, polypeptides or monospecific antibodies; methods for
CC detecting the presence of Helicobacter in samples using e.g.
CC the polypeptides or monospecific antibodies; and methods for
CC purifying the polypeptides by antibody-based affinity
CC chromatography.
XX
SQ Sequence 442 AA;

Query Match 9.6%; Score 90.5; DB 19; Length 442;
Best Local Similarity 22.1%; Pred. No. 0.027;
Matches 45; Conservative 46; Mismatches 54; Indels 59; Gaps 11;

```
Qy      18 FSFPLTMIALAISMSSWFNIWNNALSD-LGHAVKSSVAPIFNLGLAIGGILIVIVGLRNL 76
           |:: |  :  ::|:    ||   |::|   : :||::       : : |||| :|    :
Db     211 ftyslgqvffslsiglginityaavtdktqnllksti------wvvlsgililslvaglmi 264

Qy      77 YSWSRVKGSLIISMG---VFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFI--AISILDK 131
           :::    |:  :| |   :| :|  || :: | |  |||:|| |::    |   :::|:
Db     265 ftfvfeyga-nvsqgtgliftslpvvfgqm-gaigilvsilfllalafagitstvallep 322

Qy     132 S----------------W--IAVLLIIGHIAMWYLH------------------FASE- 153
           |                |  :|:: ::| : :: ||                   |||
Db     323 svmylteryqysrfkvtwglvalifvvgvvlifslhkdykdyltffekslfdwldfasst 382

Qy     154 --IPRGAAIPELLAVFSFLPFYIR 175
             :| |      :| | |: : ::
Db     383 iimplgg-----matfifmgwvlk 401
```

RESULT 4
G14660
ID G14660 standard; Protein; 258 AA.
XX
AC G14660;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14607.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%; Score 89.5; DB 21; Length 258;
Best Local Similarity 23.3%; Pred. No. 0.018;
Matches 48; Conservative 38; Mismatches 69; Indels 51; Gaps 11;

```
Qy       1 MSLNKHSWMDMIIFILSFSFPLTMIALA------------ISMSSWF--NIWNNALSDL 45
             :::: | |:   | |  :  ||     |             :||:|||   ::::   ||
Db      43 lamhfhgwlsffimiy-yklplkqdrtayyeyvglwhiygllsmnswfwsavfhsrdvdl 101

Qy      46 GHAVKSSVAPIFNLGLAIGGILIVIVGLRNL---YSWSRVKGSLIISMGVFLNL--IGVF 100
             :  | |      :|| |  :::  ||        :|| |  |   |  ::  |  :
Db     102 terldyssa------vailgfslilailrtfdirveaarvmvsapilafvtthilyinfy 155

Qy     101 DEVYGWIHFLVSVLFFLS---IIAYFIAISILDKSW-IAVLLIIGHIAMWYLHFASEIPR 156
             ||| : :| |   :|   : | : |:|    :| : |::| | :||
Db     156 kldygw-nmivcvamgvsqlflwarwaavsshpsnwklwvvviagglam---------- 203

Qy     157 GAAIPELLAVFSFLPFYIRDYFKSYT 182
               || :: | |:    || :::
Db     204 ------lleiydfppy--egyfdahs 221
```

RESULT 5
G14825
ID G14825 standard; Protein; 258 AA.
XX
AC G14825;
XX
DT 17-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 14834.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

```
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             9.5%;  Score 89.5;  DB 21;  Length 258;
  Best Local Similarity   23.3%;  Pred. No. 0.018;
  Matches   48;  Conservative   38;  Mismatches   69;  Indels   51;  Gaps   11;

Qy       1 MSLNKHSWMDMIIFILSFSFPLTMIALA----------ISMSSWF--NIWNNALSDL 45
           :::: | |:   | |  :  ||     |            :||:|||   ::::   ||
Db      43 lamhfhgwlsffimiy-yklplkqdrtayyeyvglwhiygllsmnswfwsavfhsrdvdl 101

Qy      46 GHAVKSSVAPIFNLGLAIGGILIVIVGLRNL---YSWSRVKGSLIISMGVFLNL--IGVF 100
             :  | |      :|| |  :::  ||        :||  |  |   |  ::  |  :
Db     102 terldyssa------vailgfslilailrtfdirveaarvmvsapilafvtthilyinfy 155

Qy     101 DEVYGWIHFLVSVLFFLS---IIAYFIAISILDKSW-IAVLLIIGHIAMWYLHFASEIPR 156
             ||| : :| |   :|   : | : |:|    :| : |::| | :||
Db     156 kldygw-nmivcvamgvsqlflwarwaavsshpsnwklwvvviagglam---------- 203

Qy     157 GAAIPELLAVFSFLPFYIRDYFKSYT 182
               || :: | |:    || :::
Db     204 ------lleiydfppy--egyfdahs 221


RESULT   6
G14659
ID   G14659 standard; Protein; 342 AA.
XX
AC   G14659;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 14606.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
```

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%; Score 89.5; DB 21; Length 342;
Best Local Similarity 23.3%; Pred. No. 0.025;
Matches 48; Conservative 38; Mismatches 69; Indels 51; Gaps 11;

```
Qy      1 MSLNKHSWMDMIIFILSFSFPLTMIALA------------ISMSSWF--NIWNNALSDL 45
          :::: | |:   | |  :  ||    |             :||:|||   ::::   ||
Db    127 lamhfhgwlsffimiy-yklplkqdrtayyeyvglwhiygllsmnswfwsavfhsrdvdl 185

Qy     46 GHAVKSSVAPIFNLGLAIGGILIVIVGLRNL---YSWSRVKGSLIISMGVFLNL--IGVF 100
           :  | |      :|| |  :::  ||        :||  |  |   |  ::  |  :
Db    186 terldyssa------vailgfslilailrtfdirveaarvmvsapilafvtthilyinfy 239

Qy    101 DEVYGWIHFLVSVLFFLS---IIAYFIAISILDKSW-IAVLLIIGHIAMWYLHFASEIPR 156
            ||| : :| |   :|   : | : |:|    :| : |::| | :||
Db    240 kldygw-nmivcvamgvsqlflwarwaavsshpsnwklwvvviagglam----------- 287

Qy    157 GAAIPELLAVFSFLPFYIRDYFKSYT 182
              || :: | |:    || :::
Db    288 ------lleiydfppy--egyfdahs 305
```

```
RESULT   7
G14824
ID   G14824 standard; Protein; 342 AA.
XX
AC   G14824;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 14833.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
```

```
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             9.5%;  Score 89.5;  DB 21;  Length 342;
  Best Local Similarity   23.3%;  Pred. No. 0.025;
  Matches   48;  Conservative   38;  Mismatches   69;  Indels   51;  Gaps   11;

Qy       1 MSLNKHSWMDMIIFILSFSFPLTMIALA----------ISMSSWF--NIWNNALSDL 45
           :::: | |:   | | : ||    |            :||:|||   ::::   ||
Db     127 lamhfhgwlsffimiy-yklplkqdrtayyeyvglwhiygllsmnswfwsavfhsrdvdl 185

Qy      46 GHAVKSSVAPIFNLGLAIGGILIVIVGLRNL---YSWSRVKGSLIISMGVFLNL--IGVF 100
             :  | |      :|| |  :::  ||       :||  |  |   |  ::  |  :
Db     186 terldyssa------vailgfslilailrtfdirveaarvmvsapilafvtthilyinfy 239

Qy     101 DEVYGWIHFLVSVLFFLS---IIAYFIAISILDKSW-IAVLLIIGHIAMWYLHFASEIPR 156
           ||| : :| |   :|   : | : |:|    :| : |::| | :||

Db     240 kldygw-nmivcvamgvsqlflwarwaavsshpsnwklwvvviagglam---------- 287

Qy     157 GAAIPELLAVFSFLPFYIRDYFKSYT 182
                 || :: | |:    || :::
Db     288 ------lleiydfppy--egyfdahs 305


RESULT   8
G14823
ID   G14823 standard; Protein; 356 AA.
XX
AC   G14823;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 14832.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
```

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%; Score 89.5; DB 21; Length 356;
Best Local Similarity 23.3%; Pred. No. 0.027;
Matches 48; Conservative 38; Mismatches 69; Indels 51; Gaps 11;

```
Qy      1 MSLNKHSWMDMIIFILSFSFPLTMIALA------------ISMSSWF--NIWNNALSDL 45
          :::: | |:   | |  :  ||     |             :||:|||   ::::   ||
Db    141 lamhfhgwlsffimiy-yklplkqdrtayyeyvglwhiyglsmnswfwsavfhsrdvdl 199

Qy     46 GHAVKSSVAPIFNLGLAIGGILIVIVGLRNL---YSWSRVKGSLIISMGVFLNL--IGVF 100
```

```
Db    200 terldyssa------vailgfslilailrtfdirveaarvmvsapilafvtthilyinfy 253

Qy    101 DEVYGWIHFLVSVLFFLS---IIAYFIAISILDKSW-IAVLLIIGHIAMWYLHFASEIPR 156
Db    254 kldygw-nmivcvamgvsqlflwarwaavsshpsnwklwvvviagglam---------- 301

Qy    157 GAAIPELLAVFSFLPFYIRDYFKSYT 182
Db    302 ------lleiydfppy--egyfdahs 319


RESULT   9
Y35687
ID   Y35687 standard; Protein; 485 AA.
XX
AC   Y35687;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Chlamydia pneumoniae transport polypeptide.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 1399-1400; Disclosure; 1912pp; English.
XX
CC   Y34584-Y35879 represent the proteins encoded by all the open reading
CC   frames in the complete genome (see X91990) of Chlamydia pneumoniae.
CC   C. pneumoniae causes respiratory disease such as pneumonia and
CC   bronchitis and is thought to be a contributing factor in heart
CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
CC   frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC   nucleotides sequences can also be used as immunogenic compositions,
CC   especially where the vector directs the expression of a neutralising
CC   epitope of C. pneumoniae.
XX
SQ   Sequence   485 AA;


  Query Match             9.5%;  Score 89.5;  DB 20;  Length 485;
  Best Local Similarity   22.3%;  Pred. No. 0.04;
  Matches   42;  Conservative   33;  Mismatches   68;  Indels   45;  Gaps    9;

Qy     12 IIFILSFSFPLTMIALAISMSSWF--------NIWNNALSDLGHAVKSSVAPIFNLGLAI 63
Db    156 iinvigtifkiipliifiiltafffklavfktdfw-------ghavtkaqpslgsvssql 208

Qy     64 GGILIV----IVGLRNLYSWS-RVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLS 118
Db    209 kgtmlvtlwafigiegavvmsgraknplsvgqatvlgflgclt-----iyilfsllpfgs 263

Qy    119 IIAYFIA----------ISILDKSWIAVL----LIIGHIAMW--YLHFASEIPRGAA--- 159
Db    264 lfqhqlanipnpstagvldilvgkwgevlmnvgliiavlsswlswtiivaeipfsaakng 323

Qy    160 -IPELLAV 166
Db    324 tfpeifti 331


RESULT  10
W79385
ID   W79385 standard; Protein; 249 AA.
XX
AC   W79385;
XX
DT   27-NOV-1998  (first entry)
XX
DE   Staphylococcus aureus protein of unknown function.
XX
KW   TraG protein; Lactococcus lactis; treatment; prevention;
KW   bacterial infection; Helicobacter pylori; vaccine.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "this residue is encoded by A"
XX
PN   WO9823738-A2.
XX
PD   04-JUN-1998.
XX
PF   24-NOV-1997;   97WO-US22092.
XX
PR   25-NOV-1996;   96US-0031469.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Warren RL;
XX
DR   WPI; 1998-322718/28.
DR   N-PSDB; V59895.
XX
PT   New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful
PT   for, e.g. diagnosis, prevention and treatment of bacterial
PT   infection(s)
XX
PS   Claim 5; Page 69; 114pp; English.
XX
CC   W79385-87 represent a Staphylococcus aureus WCHU (NCIMB 40771)
CC   protein of unknown function (W79385) and proteins that have
CC   homology with an Escherichia coli phophotransferase system II permease
CC   (W78386-87). The proteins are used to generate antibodies and to screen
CC   for antimicrobials. The products are used to treat or prevent bacterial
CC   infections, particularly where caused by S. aureus but also against
CC   Helicobacter pylori. Particular applications are to treat subjects before
CC   surgery or insertion of an in-dwelling device (alternatively the device
CC   itself is impregnated before placement). The nucleic acid sequence is
CC   used as sources of antisense sequences (for therapeutic use) or
CC   regulatory elements for controlling expression of bacterial genes, and
CC   for antibacterial screening. The protein can be also used as a vaccine.
XX
SQ   Sequence   249 AA;


  Query Match             9.2%;  Score 87;  DB 19;  Length 249;
  Best Local Similarity   23.7%;  Pred. No. 0.033;
  Matches   47;  Conservative   32;  Mismatches   59;  Indels   60;  Gaps   12;

Qy     12 IIFILSFSFPLTM-------IALAISMS-SWFNIWNNAL---------SDLGHAVKS--- 51
Db     43 vvfvv--glplslakkapghaalaalmgylmfntfinailtqwphtfganlekgvenvpg 100
```

```
Qy      52 --SVAPIFNLGLAI-GGILI--VIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGW 106
           |:| |  |   | |||:|  :|  : | |   |:             ::|||
Db     101 lksiagiatldtnilggiiisaiitwihnryyskrlp----------emvgvfq----g 145

Qy     107 IHFLVSVLFFLSIIAYFIAISILDKSWIAV----------LLIIGHIAMWYLHFASEIPR 156
           : |:|:: ||: :    |   |    | |          ::  |:| :|  ||   :
Db     146 ltfvvtisffvmlplaaitcvi----wptvqhgigsmqhfiiasgyigvwlyhflerv-- 199

Qy     157 GAAIPELLAVFSFLPFYI 174
           ||  |  | : |  :
Db     200 --liptglhhfiyapiev 215


RESULT  11
Y70156
ID   Y70156 standard; Protein; 412 AA.
XX
AC   Y70156;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Staphylococcus aureus protein (168339_1).
XX
KW   Vaccine; antibacterial; prevention; attenuation; detection;
KW   Staphylococcal infection; neonatal conjunctivitis; skin infection;
KW   toxic shock syndrome; osteomyelitis.
XX
OS   Staphylococcus aureus.
XX
PN   WO200012678-A2.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US19726.
XX
PR   01-SEP-1998;   98US-0098964.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey CC,  Choi GH;
XX
DR   WPI; 2000-237864/20.
DR   N-PSDB; Z51236, Z51240.
XX
PT   Staphylococcus aureus polypeptide useful for preventing or attenuating
PT   a Staphylococcal infection comprises one of 32 sequences of 100-1277
PT   amino acids or their fragments -
XX
PS   Claim 9; Page 22; 144pp; English.
XX
CC   The present sequence is a protein (168339_1)
CC   from Staphylococcus aureus genomic DNA library.
CC   This sequence is useful in preparation of vaccines for
CC   prevention or attenuation of Staphylococcal infections (especially
CC   S.aureus infections) which may cause conditions such as neonatal
CC   conjunctivitis, osteomyelitis, skin infections and toxic shock syndrome.
CC   The present sequence is also useful for detecting Staphylococcal
CC   infections in biological samples.
XX
SQ   Sequence   412 AA;


  Query Match          9.2%;  Score 87;  DB 21;  Length 412;
  Best Local Similarity   20.7%;  Pred. No. 0.063;
  Matches   25;  Conservative   30;  Mismatches   30;  Indels   36;  Gaps    5;

Qy      37 IWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNL 96
           :| ||:|     :|      : ||  :| :  | :|       |:: | |:::   :| :
Db     290 vwinais----vik------ytlgfgvglvdyvhig-------sqingillvahntylqi 332

Qy      97 IGVFDEVYG--WIHFLVSVLF----------------FLSIIAYFIAISILDKSWIAVL 137
                :  ::|  :| |:: :||                |::: ||: :|  :  ::| :
Db     333 faewgilfgalfiifmlyllfelfrfnisgknvtaivvmltmliyfltvsfnnsryvafi 392

Qy     138 L 138
           |
Db     393 l 393


RESULT  12
W98458
ID   W98458 standard; Protein; 435 AA.
XX
AC   W98458;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 710 protein.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   N-PSDB; X14177.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 8; Page 805-807; 2054pp; English.
XX
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence   435 AA;


  Query Match          9.2%;  Score 87;  DB 19;  Length 435;
  Best Local Similarity   18.2%;  Pred. No. 0.067;
  Matches   40;  Conservative   39;  Mismatches   73;  Indels   68;  Gaps    7;

Qy      22 LTMIALAISMSSWFN-IWNNALSDLGHAVKSSVAPIFNLGL--------------AIGG 65
           |:  |:   :|| || :| |  : :  :|   ::  |: ||              |: |
Db      49 lsataiitilssvfnglwan--tpiamsvglglsayfsfglvqglklpwqsalgivalsg 106

Qy      66 ILIVIVGLRNLYSW------SRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVS------- 112
            : ||:     ||      | :: ::   :| |:  ||: :     :| :|:
Db     107 aifvilsftkfrswvmrsipsdlrravsagigafiafiglke-----mhivvthxatlvt 161

Qy     113 ---------VLFFLSIIAYFIAISILDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPEL 163
                    :|  : ||  |   ::  :    : ::|  |  | |  |       ::|
Db     162 lgdfgdphvllgvvgiiltfalytlkirgsfiiavlitsilawvlklapypseffsmpas 221
```

```
Qy    164 LAVFSF----------------------LPFYIRDYFKS 180
          :   :|                      : |:: | | |
Db    222 igpiafqldfkgiffdasgaftlalvpviitffvtdlfds 261
```

RESULT 13
G14661
ID G14661 standard; Protein; 214 AA.
XX
AC G14661;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14608.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

```
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             9.2%;  Score 86.5;  DB 21;  Length 214;
  Best Local Similarity   23.9%;  Pred. No. 0.031;
  Matches   48;  Conservative   34;  Mismatches   68;  Indels   51;  Gaps   11;

Qy       6 HSWMDMIIFILSFSFPLTMIALA----------ISMSSWF--NIWNNALSDLGHAVK 50
           | |:   | |  :  ||     |            :||:|||   ::::   ||   :
Db       4 hgwlsffimiy-yklplkqdrtayyeyvglwhiygllsmnswfwsavfhsrdvdlterld 62

Qy      51 SSVAPIFNLGLAIGGILIVIVGLRNL---YSWSRVKGSLIISMGVFLNL--IGVFDEVYG 105
             | |      :|| |  :::  ||        :||  |  |   |  ::  |  :   ||
Db      63 yssa------vailgfslilailrtfdirveaarvmvsapilafvtthilyinfykldyg 116

Qy     106 WIHFLVSVLFFLS---IIAYFIAISILDKSW-IAVLLIIGHIAMWYLHFASEIPRGAAIP 161
           | : :| |   :|   : | : |:|    :| : |::| | :||
Db     117 w-nmivcvamgvsqlflwarwaavsshpsnwklwvvviagglam---------------- 159

Qy     162 ELLAVFSFLPFYIRDYFKSYT 182
             || :: | |:    || :::
Db     160 -lleiydfppy--egyfdahs 177


RESULT  14
G36489
ID   G36489 standard; Protein; 397 AA.
XX
AC   G36489;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 44725.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
```

PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.2%; Score 86.5; DB 21; Length 397;
Best Local Similarity 21.6%; Pred. No. 0.068;
Matches 30; Conservative 32; Mismatches 56; Indels 21; Gaps 5;

Qy 24 MIALAISMSSW------FNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRNLY 77

```
Db    105 vlelgllfamwylfniyfniynkqvlkalhapmtvtlvqf----avgsvlitimwvlnly 160

Qy     78 SWSRVKG---SLIISMGVFLNLIGVFDEV------YGWIHFLVSVLFFLSIIAYFIAISI 128
Db    161 krpkisgaqlaailplavvhtlgnlftnmslgkvsvsfthtikamepffsvllsamflge 220

Qy    129 LDKSWI--AVLLIIGHIAM 145
Db    221 kptpwvlgaivpivggval 239


RESULT  15
Y35302
ID   Y35302 standard; Protein; 428 AA.
XX
AC   Y35302;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Amino acid sequence of a Chlamydia pneumoniae  protein.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 1119; Disclosure; 1912pp; English.
XX
CC   Y34584-Y35879 represent the proteins encoded by all the open reading
CC   frames in the complete genome (see X91990) of Chlamydia pneumoniae.
CC   C. pneumoniae causes respiratory disease such as pneumonia and
CC   bronchitis and is thought to be a contributing factor in heart
CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
CC   frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC   nucleotides sequences can also be used as immunogenic compositions,
CC   especially where the vector directs the expression of a neutralising
CC   epitope of C. pneumoniae.
XX
SQ   Sequence   428 AA;


  Query Match             9.1%;  Score 86;  DB 20;  Length 428;
  Best Local Similarity   23.1%;  Pred. No. 0.086;
  Matches   48;  Conservative   30;  Mismatches   56;  Indels   74;  Gaps   11;

Qy      6 HSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSV---------- 53
Db      1 ysmlpliifvl-------------lcgfytswnigandvanavgpsvgsgvltlrqavv 46

Qy     54 -APIFN------LGLAIGGIL---IVIVGLRNLYSWSRVKG--SLIISMGVFLNLIGVFD 101
Db     47 iaaifeffgalllgdrvagtiessivsvtnpmiasgdymygmtaallatgvwlqlasff- 105

Qy    102 EVYGW----IHFLVSVLFFLSI------IAYFIAISILDKSWIAVLLIIGHIAMWYLHFA 151
Db    106 ---gwpvstthsivgavigfglvlgkgtiiywnsvgiiliswilspfmggcva--yl--- 157

Qy    152 SEIPRGAAIPELLAVFSFLP---FYIRD 176
Db    158 -------------ifsfirrhifyknd 171


Search completed: May 30, 2001, 15:23:49
Job time: 99 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:22:30 ; Search time 11.2 Seconds
(without alignments)
315.607 Million cell updates/sec

Title: US-09-382-242-33
Perfect score: 942
Sequence: 1 MSLNKHSWMDMIIFILSFSF.....AVFSFLPFYIRDYFKSYTKR 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	942	100.0	184	2	US-08-602-359A-33	Sequence 33, Appl
2	86	9.1	473	1	US-08-597-236-13	Sequence 13, Appl
3	86	9.1	473	1	US-08-746-682A-13	Sequence 13, Appl
4	85	9.0	279	3	US-08-549-515-7	Sequence 7, Appli
5	84.5	9.0	397	1	US-08-098-141-2	Sequence 2, Appli
6	84	8.9	363	4	US-09-046-086-2	Sequence 2, Appli
7	78.5	8.3	330	1	US-08-118-270-20	Sequence 20, Appl
8	78.5	8.3	330	5	PCT-US93-08528-20	Sequence 20, Appl
9	77	8.2	266	2	US-08-225-480-7	Sequence 7, Appli
10	77	8.2	266	2	US-09-097-759-4	Sequence 4, Appli
11	77	8.2	299	2	US-09-097-759-3	Sequence 3, Appli
12	77	8.2	306	2	US-09-097-759-2	Sequence 2, Appli
13	77	8.2	341	1	US-08-423-564-5	Sequence 5, Appli
14	76.5	8.1	394	2	US-08-902-853-1	Sequence 1, Appli
15	76.5	8.1	461	1	US-08-194-338-4	Sequence 4, Appli
16	76	8.1	330	1	US-08-118-270-19	Sequence 19, Appl
17	76	8.1	330	5	PCT-US93-08528-19	Sequence 19, Appl
18	76	8.1	355	1	US-07-759-568-1	Sequence 1, Appli
19	76	8.1	355	1	US-08-450-393A-8	Sequence 8, Appli
20	76	8.1	355	2	US-08-390-000A-5	Sequence 5, Appli
21	76	8.1	355	4	US-08-446-669-8	Sequence 8, Appli
22	76	8.1	355	5	PCT-US95-00476-8	Sequence 8, Appli
23	76	8.1	360	1	US-08-202-056-7	Sequence 7, Appli
24	75	8.0	1531	1	US-08-141-893-2	Sequence 2, Appli
25	75	8.0	1531	1	US-08-463-092B-2	Sequence 2, Appli
26	75	8.0	1531	1	US-08-463-092B-4	Sequence 4, Appli
27	75	8.0	1531	2	US-08-462-109A-2	Sequence 2, Appli.
28	75	8.0	1531	2	US-08-462-109A-4	Sequence 4, Appli
29	75	8.0	1531	2	US-08-460-907B-2	Sequence 2, Appli
30	75	8.0	1531	2	US-08-460-907B-4	Sequence 4, Appli
31	75	8.0	1531	3	US-08-463-179A-2	Sequence 2, Appli
32	75	8.0	1531	3	US-08-463-179A-4	Sequence 4, Appli
33	75	8.0	1531	3	US-08-461-384B-2	Sequence 2, Appli
34	75	8.0	1531	3	US-08-461-384B-4	Sequence 4, Appli
35	75	8.0	1531	3	US-08-407-207A-2	Sequence 2, Appli
36	74	7.9	458	1	US-08-310-271-2	Sequence 2, Appli
37	74	7.9	458	3	US-09-292-071-27	Sequence 27, Appl
38	74	7.9	458	3	US-09-292-071-29	Sequence 29, Appl
39	74	7.9	458	4	US-09-292-069A-27	Sequence 27, Appl
40	74	7.9	458	4	US-09-292-069A-29	Sequence 29, Appl
41	74	7.9	482	4	US-08-637-823B-32	Sequence 32, Appl
42	73.5	7.8	652	1	US-08-050-684-2	Sequence 2, Appli
43	73.5	7.8	652	1	US-08-582-719-2	Sequence 2, Appli
44	73	7.7	341	2	US-08-846-762-92	Sequence 92, Appl
45	73	7.7	1090	3	US-08-307-896-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT   1
US-08-602-359A-33
; Sequence 33, Application US/08602359A
; Patent No. 5942430
;  GENERAL INFORMATION:
;    APPLICANT:  ROBERTSON, Daniel E.
;    APPLICANT:  MURPHY, Dennis
;    APPLICANT:  REID, John
;    APPLICANT:  MAFFIA, Anthony
;    APPLICANT:  LINK, Steven
;    APPLICANT:  SWANSON, Ronald V.
;    APPLICANT:  WARREN, Patrick V.
;    APPLICANT:  KOSMOTKA, Anna
;    TITLE OF INVENTION:  ESTERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & RICHARDSON P.C.
;      STREET:  4225 EXECUTIVE SQUARE, STE 1400
;      CITY:  LA JOLLA
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,359A
;      FILING DATE:  February 16, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-678-5070
;      TELEFAX:  619-678-5099
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  184 AMINO ACIDS
;      TYPE:  AMINO ACID
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-602-359A-33
```

```
  Query Match           100.0%;  Score 942;  DB 2;  Length 184;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-89;
  Matches  184;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MSLNKHSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSLNKHSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLG 60

Qy     61 LAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSII 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSII 120

Qy    121 AYFIAISILDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYIRDYFKS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AYFIAISILDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYIRDYFKS 180

Qy    181 YTKR 184
          ||||
Db    181 YTKR 184


RESULT   2
US-08-597-236-13
; Sequence 13, Application US/08597236
; Patent No. 5733765
;  GENERAL INFORMATION:
;    APPLICANT:  STINGELE, Franscesca
;    APPLICANT:  MOLLET, Beat
;    TITLE OF INVENTION:  LACTIC BACTERIA PRODUCING
;    TITLE OF INVENTION:  EXOPOLYSACCHARIDES
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americans
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/597,236
;      FILING DATE:
;      CLASSIFICATION:  426
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 95201669.9
;      FILING DATE:  20-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci A., Allan
;      REGISTRATION NUMBER:  30256
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  473 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-597-236-13


  Query Match             9.1%;  Score 86;  DB 1;  Length 473;
  Best Local Similarity   24.5%;  Pred. No. 0.36;
  Matches   52;  Conservative   36;  Mismatches   74;  Indels   50;  Gaps   11;

Qy      2 SLNKHSWMDMIIFILSFSFPLT----MIALAISMSS-W-------FNIWNNALSDLGHAV 49
          |: | | :   : :  ||   |      |  |:::|: |       ||::    | |
Db     74 SMTKESVLTNSLVVWCFSAVFTCLGACIIYALNLSNKWYLALLLTFNLFQGGQSILSQYA 133

Qy     50 KS-SVAPIFNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIH 108
          :    : ||    | ||:::             : |:| |   |:| |      : |::
Db    134 RGIGKSKIF----AAGGVILTF----------LTGALNILFLVYLPL-----GITGYLM 173

Qy    109 FLV-----SVLFFLSIIAYF--IAISILDKSWI------AVLLIIGHIAMWYLHFASE-- 153
          ||     ::|||   :: :  |:  |:||  |      |: ||   |  | |: :|
Db    174 SLVLANVGTILFFAGTLSIWKEISFKIIDKKLIWQMLYYALPLIPSSILWWLLNASSRYF 233

Qy    154 --IPRGAAIPELLAVFSFLPFYIRDYFKSYTK 183
            ||    |||| : :|  |  :   :|:
Db    234 VLFFLGAGANGLLAVATKIPSIISIFNTIFTQ 265


RESULT   3
US-08-746-682A-13
; Sequence 13, Application US/08746682A
; Patent No. 5786184
;  GENERAL INFORMATION:
;    APPLICANT:  STINGELE, Franscesca
;    APPLICANT:  MOLLET, Beat
;    TITLE OF INVENTION:  LACTIC BACTERIA PRODUCING
;    TITLE OF INVENTION:  EXOPOLYSACCHARIDES
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americans
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/746,682A
;      FILING DATE:  14-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/597,236
;      FILING DATE:  20-JUN-1995
;      APPLICATION NUMBER:  EP 95201669.9
;      FILING DATE:  20-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci A., Allan
;      REGISTRATION NUMBER:  30256
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  473 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-746-682A-13


  Query Match             9.1%;  Score 86;  DB 1;  Length 473;
  Best Local Similarity   24.5%;  Pred. No. 0.36;
  Matches   52;  Conservative   36;  Mismatches   74;  Indels   50;  Gaps   11;

Qy      2 SLNKHSWMDMIIFILSFSFPLT----MIALAISMSS-W-------FNIWNNALSDLGHAV 49
          |: | | :   : :  ||   |      |  |:::|: |       ||::    | |
Db     74 SMTKESVLTNSLVVWCFSAVFTCLGACIIYALNLSNKWYLALLLTFNLFQGGQSILSQYA 133

Qy     50 KS-SVAPIFNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIH 108
          :    : ||    | ||:::             : |:| |   |:| |      : |::
Db    134 RGIGKSKIF----AAGGVILTF----------LTGALNILFLVYLPL-----GITGYLM 173

Qy    109 FLV-----SVLFFLSIIAYF--IAISILDKSWI------AVLLIIGHIAMWYLHFASE-- 153
          ||     ::|||   :: :  |:  |:||  |      |: ||   |  | |: :|
Db    174 SLVLANVGTILFFAGTLSIWKEISFKIIDKKLIWQMLYYALPLIPSSILWWLLNASSRYF 233

Qy    154 --IPRGAAIPELLAVFSFLPFYIRDYFKSYTK 183
            ||    |||| : :|  |  :   :|:
Db    234 VLFFLGAGANGLLAVATKIPSIISIFNTIFTQ 265
```

```
RESULT   4
US-08-549-515-7
; Sequence 7, Application US/08549515
; Patent No. 6054123
;  GENERAL INFORMATION:
;    APPLICANT:  Loosmore, Sheena M
;    APPLICANT:  Klein, Michel H
;    TITLE OF INVENTION:  Haemophilus Influenzae
;    TITLE OF INVENTION:  Dimethylsulphoxide Reductase Enzyme
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 Unviersity Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/549,515
;      FILING DATE:  27-OCT-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-522
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-549-515-7


  Query Match             9.0%;  Score 85;  DB 3;  Length 279;
  Best Local Similarity   19.4%;  Pred. No. 0.24;
  Matches   42;  Conservative   47;  Mismatches   67;  Indels   60;  Gaps    9;

Qy       6 HSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGG 65
           |  | :|: :|   |  ::: | : : ::     |:|: :| :: |:        :| |
Db      44 HKVMFVILALLGIGFIASIMHLGLPIRAF-----NSLNRVGSSMMSN-------EIAAGA 91

Qy      66 ILIVIVGLRNLYSWSRVKGSLIISMG----VFLNLIG-----VFDEVY------GWIHFL 110
           |   : |    |    : | : :|:|    :   |||     | ::||       | : |
Db      92 IFFTLAG---FYWLIAILGKMPVSLGNVWRIVTALIGILFMYVMNQVYHITSIPTWNNAL 148

Qy     111 VSVLFFLSIIAYFIAIS--------------------ILDKSWIAVLLIIGHIAMWYLH 149
             |  |:|:::   : :|                     :  | :|:: |     :  :|
Db     149 TSWSFYLTVVLGGLTLSYALLIPNKQREYQLQHLPSLFAIGVSLVAIVAIYQGFNLHNIH 208

Qy     150 FASE-----IPRGAAIPE----LLAVFSFLPFYIRD 176
             | :     :|  | :     ||:: :|| | :::
Db     209 SAIQNAADLVPNYAIMTVTRLCLLSIVAFLLFRVKN 244


RESULT   5
US-08-098-141-2
; Sequence 2, Application US/08098141
; Patent No. 5441875
;  GENERAL INFORMATION:
;    APPLICANT:  Hediger Ph.D., Matthias A.
;    TITLE OF INVENTION:  Urea Transporter Polypeptide
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Choate, Hall & Stewart
;      STREET:  Exchange Place, 53 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02119
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,141
;      FILING DATE:  19930723
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kaplan Esq., Warren A.
;      REGISTRATION NUMBER:  34,199
;      REFERENCE/DOCKET NUMBER:  092662-001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617:227-5020
;      TELEFAX:  617:227-7566
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  397 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-098-141-2


  Query Match             9.0%;  Score 84.5;  DB 1;  Length 397;
  Best Local Similarity   20.6%;  Pred. No. 0.42;
  Matches   47;  Conservative   32;  Mismatches   60;  Indels   89;  Gaps   13;

Qy       8 WMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGIL 67
           |: : : ::| | |:   ||    | |         ||         |:| |   |   |
Db     144 WLLLPVIVMSMSCPILSSALGTIFSKW---------DL---------PVFTLPFNIAVTL 185

Qy      68 -IVIVGLRNLY---------------SWSRVKGSLI---ISMGVFLNLIGVFDEVYG--- 105
            :   |  ||:               :|| ::  |:   | :|:         :|||
Db     186 YLAATGHYNLFFPTTLLQPVSSVPNITWSEIQVPLLLRAIPVGI--------GQVYGCDN 237

Qy     106 -WIHFLVSVLFFL------------SIIAYFIAISILD-----------KSWIAVLLII 140
            |   :  :  |:             | :    |::|               | :| : :
Db     238 PWTGGIFLIALFISSPLICLHAAIGSTMGMLAALTIATPFDSIYFGLCGFNSTLACIAVG 297

Qy     141 G--HIAMWYLH--------FASEIPRGAAIPELLAVFSF----LPFYI 174
           |  ::  |  |        ||: :  |||:  :|:||       || |
Db     298 GMFYVITWQTHLLAVACALFAAYV--GAALTNVLSVFGLPTCTWPFCI 343


RESULT   6
US-09-046-086-2
; Sequence 2, Application US/09046086
; Patent No. 6127147
;  GENERAL INFORMATION:
;    APPLICANT:  Wallis, Nicola G.
;    TITLE OF INVENTION:  No. 6127147el Compounds
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert, Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2793
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/09/046,086
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/043,489
;      FILING DATE:  10-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Falk, Stephen T
;      REGISTRATION NUMBER:  36,795
;      REFERENCE/DOCKET NUMBER:  GM50025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2488
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-046-086-2


  Query Match             8.9%;  Score 84;  DB 4;  Length 363;
  Best Local Similarity   27.3%;  Pred. No. 0.42;
  Matches   30;  Conservative   21;  Mismatches   45;  Indels   14;  Gaps    4;

Qy      82 VKGSLIISMGVFLNLI----GV-FDEVYG--WIHFLVSVLFFLSIIAYFIAISILDKSWI 134
           :| : |  :   | ||    |: |:||||  |:: :  ::| || :   |  : |:
Db       4 LKDTSIAEISSILYLIFPIAGIFFNEVYGPKWLYIISVIVFSLSYLILVIVNNRLNTLMF 63

Qy     135 AVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYIRDYFKSYTKR 184
           :|||| :  : |  |:         | |   | : | :   ||:  |:
Db      64 YILLIIHYFIICYFVFSVH-------PMLSLFFFYSAFAVPFTFKNNVKK 106


RESULT   7
US-08-118-270-20
; Sequence 20, Application US/08118270
; Patent No. 5508384
;  GENERAL INFORMATION:
;    APPLICANT:  Murphy, Randall B.
;    APPLICANT:  Schuster, David I.
;    TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
;    TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;    NUMBER OF SEQUENCES:  348
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/118,270
;      FILING DATE:  09-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/943,236
;      FILING DATE:  10-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Townsend, Kevin G.
;      REGISTRATION NUMBER:  34,033
;      REFERENCE/DOCKET NUMBER:  MURPHY=2A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  330 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-118-270-20


  Query Match             8.3%;  Score 78.5;  DB 1;  Length 330;
  Best Local Similarity   28.9%;  Pred. No. 1.4;
  Matches   33;  Conservative   17;  Mismatches   43;  Indels   21;  Gaps    5;

Qy      25 IALAISMSSWFNIWNNALSDLGHAVKSSV-AP--IFNLGLAIGGILIVIVGLRNLYSWSR 81
           || ||:    | |: |||  :      |: ||  :| : :|   ||:
Db       4 IAAAITFLILFTIFGNALVIIAVLTSRSLRAPQNLFLVSIAAADILV---------- 50

Qy      82 VKGSLIISMGVFLNLIG--VFDEVYGWIHFLVSVLFFLSIIAYFIAISILDKSW 133
           :|||   :   |:|   |   :  ::  : |||  | | :  ||| ||: |
Db      51 --ATLIIPFSLANELLGYWYFRRTWCEVYLALDVLFCTSSIVHLCAIS-LDRYW 101


RESULT   8
PCT-US93-08528-20
; Sequence 20, Application PC/TUS9308528
;  GENERAL INFORMATION:
;    APPLICANT:  New York University
;    TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
;    TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;    NUMBER OF SEQUENCES:  348
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08528
;      FILING DATE:  09-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/943,236
;      FILING DATE:  10-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Townsend, Kevin G.
;      REGISTRATION NUMBER:  34,033
;      REFERENCE/DOCKET NUMBER:  MURPHY=2 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  330 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-08528-20


  Query Match             8.3%;  Score 78.5;  DB 5;  Length 330;
  Best Local Similarity   28.9%;  Pred. No. 1.4;
  Matches   33;  Conservative   17;  Mismatches   43;  Indels   21;  Gaps    5;

Qy      25 IALAISMSSWFNIWNNALSDLGHAVKSSV-AP--IFNLGLAIGGILIVIVGLRNLYSWSR 81
           || ||:    | |: |||  :      |: ||  :| : :|   ||:
Db       4 IAAAITFLILFTIFGNALVIIAVLTSRSLRAPQNLFLVSIAAADILV---------- 50
```

```
Qy      82 VKGSLIISMGVFLNLIG--VFDEVYGWIHFLVSVLFFLSIIAYFIAISILDKSW 133
           :|||   :   |:|   |   :  ::  : |||  | | :  ||| ||: |
Db      51 --ATLIIPFSLANELLGYWYFRRTWCEVYLALDVLFCTSSIVHLCAIS-LDRYW 101


RESULT   9
US-08-225-480-7
; Sequence 7, Application US/08225480
; Patent No. 5817793
;  GENERAL INFORMATION:
;    APPLICANT:  LEVY, STUART B.
;    TITLE OF INVENTION:  MULTIPLE ANTIBIOTIC RESISTANCE OPERON
;    TITLE OF INVENTION:  ASSAYS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  WOLF, GREENFIELD & SACKS, P.C.
;      STREET:  600 ATLANTIC AVENUE
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/225,480
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,085
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GATES, EDWARD R.
;      REGISTRATION NUMBER:  31,616
;      REFERENCE/DOCKET NUMBER:  T0359/7003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  266 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-225-480-7


  Query Match             8.2%;  Score 77;  DB 2;  Length 266;
  Best Local Similarity   24.1%;  Pred. No. 1.5;
  Matches   38;  Conservative   24;  Mismatches   66;  Indels   30;  Gaps    6;

Qy      16 LSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRN 75
           ::|  |  : :| :   ::| |   | :  |  :      :  : ||| |:|::|    |
Db      47 INFGMPAGLASLVLQAQAFFTIMLGAFT-FGERLHGK--QLAGIALAIFGVLVLIEDSLN 103

Qy      76 LYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIH-------------FLVSVLFFLSIIA 121
                 |  :  :|  | |   |    | |                ||     : ||
Db     104 --------GQHVAMLGFMLTLAAAFSWACGNIFNKKIMSHSTRPAVMSLVIWSALIPIIP 155

Qy     122 YFIAISILDKS--WIAVLLIIGH---IAMWYLHFASEI 154
           :|:|  ||| |   |  |: |    ::: || | : |
Db     156 FFVASLILDGSATMIHSLVTIDMTTILSLMYLAFVATI 193


RESULT  10
US-09-097-759-4
; Sequence 4, Application US/09097759A
; Patent No. 5972663
; GENERAL INFORMATION:
;  APPLICANT: Winterhalter Mr., Christopher
;  APPLICANT: Leinfelder Mr., Walfred
;  TITLE OF INVENTION: Microorganisms and Processes  for the Fermentative
;  TITLE OF INVENTION: Preparation of L-cysteine,
;  TITLE OF INVENTION: L-cysteine, N-acetylserine or Thiazolidine Derivatives
;  FILE REFERENCE: Winterhalter
;  CURRENT APPLICATION NUMBER: US/09/097,759A
;  CURRENT FILING DATE: 1998-06-16
;  EARLIER APPLICATION NUMBER: DE 197 26 083
;  EARLIER FILING DATE: 1997-06-19
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 266
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-097-759-4


  Query Match             8.2%;  Score 77;  DB 2;  Length 266;
  Best Local Similarity   24.1%;  Pred. No. 1.5;
  Matches   38;  Conservative   24;  Mismatches   66;  Indels   30;  Gaps    6;

Qy      16 LSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRN 75
           ::|  |  : :| :   ::| |   | :  |  :      :  : ||| |:|::|    |
Db      47 INFGMPAGLASLVLQAQAFFTIMLGAFT-FGERLHGK--QLAGIALAIFGVLVLIEDSLN 103

Qy      76 LYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIH-------------FLVSVLFFLSIIA 121
                 |  :  :|  | |   |    | |                ||     : ||
Db     104 --------GQHVAMLGFMLTLAAAFSWACGNIFNKKIMSHSTRPAVMSLVIWSALIPIIP 155

Qy     122 YFIAISILDKS--WIAVLLIIGH---IAMWYLHFASEI 154
           :|:|  ||| |   |  |: |    ::: || | : |
Db     156 FFVASLILDGSATMIHSLVTIDMTTILSLMYLAFVATI 193


RESULT  11
US-09-097-759-3
; Sequence 3, Application US/09097759A
; Patent No. 5972663
; GENERAL INFORMATION:
;  APPLICANT: Winterhalter Mr., Christopher
;  APPLICANT: Leinfelder Mr., Walfred
;  TITLE OF INVENTION: Microorganisms and Processes  for the Fermentative
;  TITLE OF INVENTION: Preparation of L-cysteine,
;  TITLE OF INVENTION: L-cysteine, N-acetylserine or Thiazolidine Derivatives
;  FILE REFERENCE: Winterhalter
;  CURRENT APPLICATION NUMBER: US/09/097,759A
;  CURRENT FILING DATE: 1998-06-16
;  EARLIER APPLICATION NUMBER: DE 197 26 083
;  EARLIER FILING DATE: 1997-06-19
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 299
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-097-759-3


  Query Match             8.2%;  Score 77;  DB 2;  Length 299;
  Best Local Similarity   24.1%;  Pred. No. 1.7;
  Matches   38;  Conservative   24;  Mismatches   66;  Indels   30;  Gaps    6;

Qy      16 LSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRN 75
           ::|  |  : :| :   ::| |   | :  |  :      :  : ||| |:|::|    |
Db      80 INFGMPAGLASLVLQAQAFFTIMLGAFT-FGERLHGK--QLAGIALAIFGVLVLIEDSLN 136

Qy      76 LYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIH-------------FLVSVLFFLSIIA 121
                 |  :  :|  | |   |    | |                ||     : ||
Db     137 --------GQHVAMLGFMLTLAAAFSWACGNIFNKKIMSHSTRPAVMSLVIWSALIPIIP 188

Qy     122 YFIAISILDKS--WIAVLLIIGH---IAMWYLHFASEI 154
           :|:|  ||| |   |  |: |    ::: || | : |
```

```
Db    189 FFVASLILDGSATMIHSLVTIDMTTILSLMYLAFVATI 226


RESULT  12
US-09-097-759-2
; Sequence 2, Application US/09097759A
; Patent No. 5972663
; GENERAL INFORMATION:
;  APPLICANT: Winterhalter Mr., Christopher
;  APPLICANT: Leinfelder Mr., Walfred
;  TITLE OF INVENTION: Microorganisms and Processes  for the Fermentative
;  TITLE OF INVENTION: Preparation of L-cysteine,
;  TITLE OF INVENTION: L-cysteine, N-acetylserine or Thiazolidine Derivatives
;  FILE REFERENCE: Winterhalter
;  CURRENT APPLICATION NUMBER: US/09/097,759A
;  CURRENT FILING DATE: 1998-06-16
;  EARLIER APPLICATION NUMBER: DE 197 26 083
;  EARLIER FILING DATE: 1997-06-19
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 306
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-097-759-2


  Query Match             8.2%;  Score 77;  DB 2;  Length 306;
  Best Local Similarity   24.1%;  Pred. No. 1.8;
  Matches   38;  Conservative   24;  Mismatches   66;  Indels   30;  Gaps    6;

Qy      16 LSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRN 75
           ::|  |  : :| :   ::| |   | :  |  :      :  : ||| |:|::|    |
Db      87 INFGMPAGLASLVLQAQAFFTIMLGAFT-FGERLHGK--QLAGIALAIFGVLVLIEDSLN 143

Qy      76 LYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIH-------------FLVSVLFFLSIIA 121
                 | :  :|  | |   |    | |               ||     : ||
Db     144 --------GQHVAMLGFMLTLAAAFSWACGNIFNKKIMSHSTRPAVMSLVIWSALIPIIP 195

Qy     122 YFIAISILDKS--WIAVLLIIGH---IAMWYLHFASEI 154
           :|:|  ||| |   |  |: |     ::: || | : |
Db     196 FFVASLILDGSATMIHSLVTIDMTTILSLMYLAFVATI 233


RESULT  13
US-08-423-564-5
; Sequence 5, Application US/08423564
; Patent No. 5654417
; Patent No. 5654417 5654417
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, Phillip I.
;    APPLICANT:  Bilge, Sima S.
;    APPLICANT:  Vary, Jr., James C.
;    TITLE OF INVENTION:  NUCLEIC ACID PROBES FOR DETECTING E. coli O157:H7
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christiansen, O'Connor, Johnson, Kindness
;      STREET:  1420 5th Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/423,564
;      FILING DATE:  14-APR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Broderick, Thomas F.
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  CHOR-1-8116
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682-8100
;      TELEFAX:  (206) 224-0779
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  341 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;      DESCRIPTION:  amino acids encoded by
;      DESCRIPTION:  nucleotides 1-1026 of SEQ ID NO:1
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
US-08-423-564-5


  Query Match             8.2%;  Score 77;  DB 1;  Length 341;
  Best Local Similarity   23.5%;  Pred. No. 2;
  Matches   43;  Conservative   28;  Mismatches   68;  Indels   44;  Gaps    9;

Qy      19 SFPLTMIALAIS-MSSW----FNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGL 73
           || :| |    | |  |       |  |  ::| |::|  |  || ||:   |:  | :|:
Db       1 SFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKVFVI-IFGLGIYSNGVQKVYMGI 59

Qy      74 RNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIH-----FLVSVLFFLSIIAYFIAISI 128
           :  |  | :  :: | : :   :|          :|      :|| |    |   :: |::
Db      60 QKAYI-SNIVNAIFILLSIITLVIS------SKLHAGLPVLIVSTLGIQYISGIYLTINL 112

Qy     129 LDKSWIA---------------------VLLIIGHIAMWY-LHFASEIPRGAAIPELLA 165
           : |  |                      :| :| :| |   :|   |  |      :|
Db     113 IIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSGDNFIISITLGVT---YVA 169

Qy     166 VFS 168
           |||
Db     170 VFS 172


RESULT  14
US-08-902-853-1
; Sequence 1, Application US/08902853
; Patent No. 5945330
;  GENERAL INFORMATION:
;    APPLICANT:  HIllman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Lal, Preeti
;    TITLE OF INVENTION:  HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/902,853
;      FILING DATE:  Herewith
;      CLASSIFICATION:  ?
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0345 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  394 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  LIVRTUT04
;       CLONE:  2516821
US-08-902-853-1


  Query Match             8.1%;  Score 76.5;  DB 2;  Length 394;
  Best Local Similarity   21.9%;  Pred. No. 2.7;
  Matches   40;  Conservative   25;  Mismatches   37;  Indels   81;  Gaps    9;

Qy      56 IFNLGLAIGGILIVIVGLRNLY----------SWSR-----VKGSLIISMGVFLNLIGVF 100
           :| |   :||: :       ||           | |     :| ||
Db     141 LFYLSSFVGGLSV-------LYHESWLWAPVMCWDRYPNQTLKPSL-------------- 179

Qy     101 DEVYGWIHFLVSVLFFLS----------------IIAYFIAISILDKSWIAVLLIIGHI 143
             | |  :|: : |:||                :| :|:|: ::  |: | || || :
Db     180 ---YWW--YLLELGFYLSLLIRLPFDVKRKDFKEQVIHHFVAVILMTFSYSANLLRIGSL 234

Qy     144 AMWYLHFASEIPRGA-----------AIPELLAVFSFLPFYIR----------DYFKSY 181
           :  || :|:    |              |  :|||: || |            |::|
Db     235 VL-LLHDSSDYLLEACKMVNYMQYQQVCDALFLIFSFVFFYTRLVLFPTQILYTTYYESI 293

Qy     182 TKR 184
           : |
Db     294 SNR 296


RESULT  15
US-08-194-338-4
; Sequence 4, Application US/08194338
; Patent No. 5474898
;  GENERAL INFORMATION:
;    APPLICANT:  Venter, John C.
;    APPLICANT:  Fraser, Claire M.
;    APPLICANT:  McCombie, William R.
;    TITLE OF INVENTION:  OCTOPAMINE RECEPTOR
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, olson and Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,338
;      FILING DATE:  08-FEB-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/676,174
;      FILING DATE:  28-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH101.001DV1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  461 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
US-08-194-338-4


  Query Match             8.1%;  Score 76.5;  DB 1;  Length 461;
  Best Local Similarity   24.3%;  Pred. No. 3.3;
  Matches   33;  Conservative   20;  Mismatches   46;  Indels   37;  Gaps    4;

Qy      63 IGGILIVIVGL--------RNLYSWSRVKGSLII-------SMGVFLNLIGVFDEVYGWI 107
           :| :|:||  |         :||:  |     :::       |:   |     | :|:  :
Db      67 VGNVLVVIAVLTSRALRAPQNLFLVSLASADILVATLVMPFSLANELMAYWYFGQVWCGV 126

Qy     108 HFLVSVLFFLSIIAYFIAISILDKSW--------------------IAVLLIIGHIAMW 146
           :  : |||  | | :  ||| ||: |                      :|| :|   |:
Db     127 YLALDVLFCTSSIVHLCAIS-LDRYWSVTGAVEYNLKRTPRRVKATIVAVWIISAVISFP 185

Qy     147 YLHFASEIPRGAAIPE 162
           |     | ||| |:
Db     186 PLVSLYRQPDGAAYPQ 201


Search completed: May 30, 2001, 15:24:03
Job time: 93 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:22:50 ; Search time 14.23 Seconds
(without alignments)
888.617 Million cell updates/sec

Title: US-09-382-242-33
Perfect score: 942
Sequence: 1 MSLNKHSWMDMIIFILSFSF.....AVFSFLPFYIRDYFKSYTKR 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	163.5	17.4	175	2	D75214	hypothetical prote
2	144.5	15.3	156	2	A71189	hypothetical prote
3	142	15.1	192	2	G71450	hypothetical prote
4	129.5	13.7	183	2	B71040	hypothetical prote
5	126.5	13.4	156	2	A75149	hypothetical prote
6	118	12.5	165	2	G72486	hypothetical prote
7	105.5	11.2	401	2	G82210	amino acid ABC tra
8	104.5	11.1	373	2	B75542	conserved hypothet
9	103.5	11.0	600	2	A82043	inner membrane cop
10	100.5	10.7	404	1	LABECA	latent membrane pr
11	98.5	10.5	360	2	E82442	probable peptide A
12	97.5	10.4	280	2	A71694	hypothetical prote
13	97.5	10.4	360	2	C82445	hypothetical prote
14	97	10.3	144	2	S26026	NADH dehydrogenase
15	96.5	10.2	485	2	H82822	NADH-ubiquinone ox
16	96	10.2	419	2	B69819	conserved hypothet
17	95	10.1	1321	2	A60165	sodium channel pro
18	94	10.0	394	2	S48522	cell division cont
19	94	10.0	510	2	E71695	hypothetical prote
20	93.5	9.9	460	2	C71884	hypothetical prote
21	93.5	9.9	870	2	T25338	hypothetical prote
22	93	9.9	383	2	S53379	probable membrane
23	93	9.9	487	2	E70480	probable sodium/pr
24	92.5	9.8	303	2	E69585	sugar transport am
25	92	9.8	407	2	E70309	hypothetical prote
26	91.5	9.7	452	2	H71706	ampg protein (ampG
27	91	9.7	313	2	T11272	NADH dehydrogenase
28	91	9.7	435	2	B71850	hypothetical prote
29	91	9.7	643	2	E69373	conserved hypothet
30	90.5	9.6	442	2	A64582	sodium- and chlori
31	90.5	9.6	532	2	S66087	amino acid transpo
32	90	9.6	274	2	S42170	NADH dehydrogenase
33	90	9.6	381	2	S24611	latent membrane pr
34	89.5	9.5	234	2	B83807	hypothetical prote
35	89.5	9.5	258	2	E72319	flagellar biosynth
36	89.5	9.5	442	2	F71930	probable transport
37	89.5	9.5	461	1	E64630	virulence factor m
38	89.5	9.5	485	2	A72006	arginine/ornithine
39	89.5	9.5	587	2	S65307	probable membrane
40	89	9.4	226	2	A42622	H+-transporting AT
41	89	9.4	243	2	S73698	hypothetical prote
42	89	9.4	253	2	C82517	conserved hypothet
43	89	9.4	277	2	B64329	hypothetical prote
44	89	9.4	404	2	H83249	sodium/glutamate s
45	88.5	9.4	294	2	T29272	hypothetical prote

ALIGNMENTS

RESULT 1
D75214
hypothetical protein PAB0162 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: D75214
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: D75214
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-175 <KAW>
A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49163.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0162
C;Superfamily: Pyrococcus abyssi hypothetical protein PAB0162

Query Match 17.4%; Score 163.5; DB 2; Length 175;
Best Local Similarity 34.8%; Pred. No. 8e-07;
Matches 55; Conservative 28; Mismatches 64; Indels 11; Gaps 5;

```
Qy      24 MIALA------ISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRNLY 77
           :||||      :    |::|  ||:|||| |      : | || |||:|::  |   |
Db      10 IIALAGVVFAYLLNRDWWSITENAISDLGR-VGLPNNWVMNCGLIIGGLLLLGYGAWRLK 68

Qy      78 SWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSIIA--YFIAISILDKSWIA 135
              :: :   ||| ||||| |   | |: ||  ||:|: |    |:|||| || :
Db      69 KSKNPGWAIYVVGSVFLVLIGVFPEGTEW-HYEVSWGFFVSMFAAMLIISISILKKSRLG 127

Qy     136 VL-LIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPF 172
           | | :  :::    |: :   | |: | ::: :|| |
Db     128 FLGLAVFCVSLPLALFSLKAFSGVAVAETVSILAFLIF 165
```

RESULT 2
A71189
hypothetical protein PH1785 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: A71189
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137
A;Accession: A71189
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

```
A;Residues: 1-156 <KAW>
A;Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA30904.1; PID:g3258221
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1785


  Query Match            15.3%;  Score 144.5;  DB 2;  Length 156;
  Best Local Similarity  31.3%;  Pred. No. 2.7e-05;
  Matches   52;  Conservative   30;  Mismatches   59;  Indels   25;  Gaps    8;

Qy      12 IIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNL---GLAIGGILI 68
           |: :: |   :  :|| || : ||:   |||||:| | |:    :|||   || | ||::
Db       7 IVMVIYF---IIGLALVISQNPWFSFTKNALSDMGSA-KNPKGWMFNLYIIGLGIIGIIV 62

Qy      69 VIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISI 128
             :  |||         | ||| : | |:|:| |     |   ::| :|     || : :
Db      63 ANLLNRNL---------LKISM-ILLILVGIFPEEEP-PHTPSAILMYL---LSFIDMGL 108

Qy     129 LDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYI 174
             ::|  :  |   | :  : |      | |||||:   : | : :
Db     109 YSRTWRIIATITFTIMVILVSFKI----GLAIPELIGAAAILSYIL 150


RESULT   3
G71450
hypothetical protein PH0262 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: G71450
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: G71450
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-192 <KAW>
A;Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29334.1; PID:g3256651
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0262
C;Superfamily: Pyrococcus abyssi hypothetical protein PAB0162


  Query Match            15.1%;  Score 142;  DB 2;  Length 192;
  Best Local Similarity  30.7%;  Pred. No. 5.2e-05;
  Matches   47;  Conservative   32;  Mismatches   44;  Indels   30;  Gaps    9;

Qy      34 WFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRNLYSW----SRVKGSLIIS 89
           |: |  ||:|||  :      : |: | :| : ::| |     :|    |:  | |:
Db      44 WWGITENAISDLGR-IGLPYNWVMNISLILGSVCLIIYG-----AWRFKKSKDIGWLLYM 97

Qy      90 MG-VFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISILDKSWI----------AVLL 138
           :| ||| |||:| |    :|: ||  ||:|:   |:|| :|  |::          :|
Db      98 LGSVFLGLIGIFPEGTN-LHYEVSWGFFVSM---FLAILLLSISFLIRGNKLGIVGLMLF 153

Qy     139 IIG-HIAMWYLHFASEIPRGAAIPELLAVFSFL 170
           ::|  :|:| |    :   | |: | :::  ||
Db     154 LLGVPLALWSL----KKFEGVAVAETISIVVFL 182


RESULT   4
B71040
hypothetical protein PH1610 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: B71040
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137
A;Accession: B71040
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-183 <KAW>
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30722.1; PID:g3258039
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH1610
C;Superfamily: Pyrococcus abyssi hypothetical protein PAB0162


  Query Match            13.7%;  Score 129.5;  DB 2;  Length 183;
  Best Local Similarity  29.4%;  Pred. No. 0.00054;
  Matches   40;  Conservative   22;  Mismatches   47;  Indels   27;  Gaps    4;

Qy      28 AISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILI------VIVGLRNLYSWSR 81
           :|| : ||: : |||||||   |:|   |:| || :    :      :|:  :|
Db      30 SISNNPWFSFFKNALSDLGDPAKASHPWIYNYGLIVTSPFVLAFSIYLIIAAKN--KIQT 87

Qy      82 VKGSLIISMGVFLNLIGVFD---------EVYGWIHFLVSVLFF----------LSIIAY 122
           | |: |    :|| |||:|            | :| | : :| :           |||  :
Db      88 VGGAYISISSLFLALIGIFHAGTRPHVFVSTYFFIQFFLGMLIYGVGSKTAIRLTSIIIF 147

Qy     123 FIAISILDKSWIAVLL 138
           :|:  |   | :| |
Db     148 ILALLGLFIPWPSVAL 163


RESULT   5
A75149
hypothetical protein PAB0225 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: A75149
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: A75149
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-156 <KAW>
A;Cross-references: GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49272.1; PID:e151
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0225


  Query Match            13.4%;  Score 126.5;  DB 2;  Length 156;
  Best Local Similarity  25.3%;  Pred. No. 0.00082;
  Matches   42;  Conservative   33;  Mismatches   64;  Indels   27;  Gaps    5;

Qy      17 SFSFPLTM-----IALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIV 71
           : |  ||:     :|| :  :|||:   |||||:| ::::    :||       : |: :
Db       4 ALSIALTLYFILGLALVVHENSWFSFSKNALSDMG-SLRNPKGWMFN-------VFIIGL 55

Qy      72 GLRNLYSWSRVKGSLIISMGVFLNLIGVFDE---VYGWIHFLVSVLFFLSIIAYFIAISI 128
           ||  | :   :|  ::     || |:||| |   ::     |  :| |  :: |
Db      56 GLLGLVTSLMLKRKILTLSMAFLVLVGVFPEEKPLHTPSAILTYILAFTDMVIYG----- 110

Qy     129 LDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYI 174
                :  ::|      : |   :  | |||||:   | | : :
Db     111 ------GIWRVVGIGTFMVMLFLINLKVGLAIPELIGAASILAYIL 150


RESULT   6
G72486
hypothetical protein APE2535 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
```

```
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: G72486
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: G72486
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-165 <KAW>
A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81551.1; PID:g5106240
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE2535
C;Superfamily: Pyrococcus abyssi hypothetical protein PAB0162


  Query Match             12.5%;  Score 118;  DB 2;  Length 165;
  Best Local Similarity   30.4%;  Pred. No. 0.0043;
  Matches   41;  Conservative   20;  Mismatches   52;  Indels   22;  Gaps    5;

Qy      32 SSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGL---RNLYSWSRVKGS-LI 87
           | ||    :| |||| ||: :  |:| || : |:|: :  |   |          || |:
Db       3 SDWFVFTEHAFSDLG-GVKARMPWIYNYGLILVGLLVTLYSLCPYRAAVERLEAFGSGLL 61

Qy      88 ISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISI---------------LDK 131
           :: |:|| ||||:       |  ||  ||| :    :|:|:                |
Db      62 LTAGLFLALIGVY-PAGTRPHVFVSTWFFLQMDMALVALSLGAYRACRNLANMAALALSL 120

Qy     132 SWIAVLLIIGHIAMW 146
           :   | ::||    |
Db     121 AAFPVYILIGITLGW 135


RESULT   7
G82210
amino acid ABC transporter, permease protein VC1361 [imported] - Vibrio cholerae (group
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C;Accession: G82210
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: G82210
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401 <HEI>
A;Cross-references: GB:AE004215; GB:AE003852; NID:g9655842; PIDN:AAF94519.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC1361
A;Map position: 1


  Query Match             11.2%;  Score 105.5;  DB 2;  Length 401;
  Best Local Similarity   20.1%;  Pred. No. 0.1;
  Matches   46;  Conservative   38;  Mismatches   84;  Indels   61;  Gaps    7;

Qy       6 HSWMDMIIFILSF-SFPLTMIALAISMSSWFNIWNNALSDL---------------GHA 48
           ||    :::  :| |    ::|: :     : | |||::|               |
Db      15 HSKHTNLLYNPTFRSVVFQLLAVCVLAFFLYTIVNNALTNLESRGIATGFAFLEQTAGFG 74

Qy      49 VKSSVAP-------------------IFNLGLAIGGILIVIVGLRNLYSWSRVKGSLII 88
           :  |: |                   :  ||: :  ||  ::|:  | |   :     :
Db      75 ISQSLIPYDETYSYGRTFLVGLLNTALVSVLGILLATILGFLIGIARLSSNWLISRLAAV 134

Qy      89 SMGVFLNLIGVFDEVYGWIHFLVSVL------------FFLSIIAYFIAISILDKS--WI 134
             :  | ||  :  ::: |   ::  |            |:|:|   ::|  | :    |:
Db     135 YIETFRNL-PLLLQIFFWYFVVLQALPSARESMHLGEWFYLNIKGLYLAKPIFESGSIWV 193

Qy     135 AVLLIIGHIAMWYLHF---------ASEIPRGAAIPELLAVFSFLPFYI 174
           | || | || | |             : | |  |  |   |  | :::
Db     194 LVALIAGIIACWVLSIWATNRQRLTGQQTPMGRYILLLCVTFPVLVYFL 242


RESULT   8
B75542
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: B75542
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: B75542
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-373 <WHI>
A;Cross-references: GB:AE001886; GB:AE000513; NID:g6457921; PIDN:AAF09838.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0252
A;Map position: 1
C;Superfamily: Bacillus subtilis conserved hypothetical protein yueF


  Query Match             11.1%;  Score 104.5;  DB 2;  Length 373;
  Best Local Similarity   27.4%;  Pred. No. 0.12;
  Matches   43;  Conservative   29;  Mismatches   44;  Indels   41;  Gaps    9;

Qy      30 SMSSWF-NIWNNALSDLGHAVKSSVAPIFNLGLAIGGI-----LIVIVGLRNLYSWSRVK 83
           ::  |  || || :  | :|: ||   :||  ::|||:     ||::: :  :  ::||
Db     142 ALRDWIANIGNNIVPILQNAL-SSTGTLFNRLISIGGVIGQVLLILLLSVYLMLDYARVN 200

Qy      84 GSLIISMG--------VFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISILDKSWIA 135
           ||: :           | :|||    | |::   :        :|| || | :    |:
Db     201 ASLLRAFPRPWQPKVLEFSDLIGT--AVGGYVRGQL-------LIAAFIGIFV----WLG 247

Qy     136 VLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPF 172
           | |:|            ||  |||  |   |: :|:
Db     248 -LTIVG------------IPSAAAIGFLAGAFNIVPY 271


RESULT   9
A82043
inner membrane copper tolerance protein cycZ VC2701 [similarity] - Vibrio cholerae (g
N;Alternate names: thiol:disulfide interchange protein DsbD
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C;Accession: A82043
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: A82043
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-600 <HEI>
A;Cross-references: GB:AE004336; GB:AE003852; NID:g9657296; PIDN:AAF95841.1; GSPDB:GN
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC2701
A;Map position: 1


  Query Match             11.0%;  Score 103.5;  DB 2;  Length 600;
  Best Local Similarity   22.5%;  Pred. No. 0.22;
```

```
  Matches   42;  Conservative   40;  Mismatches   82;  Indels   23;  Gaps    7;

Qy       5 KHSWMDMIIFILSFSFPLTMIAL-AISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAI 63
           :| :: | : :|  :  |:|  | :: : |    | |:|||:       |:  :| :|
Db     271 QHPYVLMGLSVLFVALALSMFGLYSLQLPSGVQTWLNSLSNAQQG--GSLPGVFAMGAIS 328

Qy      64 G------------GILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVF-----DEVYGW 106
           |            | |: :    :| : :    :| | ||: | |: ||      :   |
Db     329 GLVCSPCTTAPLSGALLYVAQSGDLLTGAVALYALAIGMGIPLILVAVFGNKLLPKAGNW 388

Qy     107 IHFLVSVLFFLSIIA-YFIAISILDKSWIAVL-LIIGHIAM-WYLHFASEIPRGAAIPEL 163
           :  : ::  |: : |  |:   |: : | :||   :|  |  |  |  : :| |     |
Db     389 MERVKTLFGFVLLAAPIFLLERIVPEFWSSVLWSALGLAAFGWLYHVKNSLPFGGWKQSL 448

Qy     164 LAVFSFL 170
           : : : |
Db     449 IGIVAIL 455


RESULT  10
LABECA
latent membrane protein LMP1 - human herpesvirus 4 (strain CAO)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C;Accession: JQ1434; G00065; S21660
R;Hu, L.F.; Zabarovsky, E.R.; Chen, F.; Cao, S.L.; Ernberg, I.; Klein, G.; Winberg, G.
J. Gen. Virol. 72, 2399-2409, 1991
A;Title: Isolation and sequencing of the Epstein-Barr virus BNLF-1 gene (LMP1) from a Ch
A;Reference number: JQ1434; MUID:92013956
A;Accession: JQ1434
A;Molecule type: DNA
A;Residues: 1-404 <HUL>
A;Cross-references: EMBL:X58140; NID:g22937; PIDN:CAA41148.1; PID:g22938
A;Note: the authors translated the codon AAA for residue 358 as Ala
A;Note: warning: the organism was shown incorrectly as Homo sapiens in GenBank accession
C;Comment: Unlike Epstein-Barr nuclear antigen 1 (EBNA-1) (see PIR:S42440), which is exp
y of nasopharyngeal carcinoma tumor biopsies.
C;Genetics:
A;Gene: LMP1; BNLF-1
A;Introns: 90/1; 119/1
C;Superfamily: Epstein-Barr virus latent membrane protein
C;Keywords: oncogene; tandem repeat; transmembrane protein
F;25-45/Domain: transmembrane #status predicted <TM1>
F;51-71/Domain: transmembrane #status predicted <TM2>
F;76-97/Domain: transmembrane #status predicted <TM3>
F;104-124/Domain: transmembrane #status predicted <TM4>
F;140-160/Domain: transmembrane #status predicted <TM5>
F;166-186/Domain: transmembrane #status predicted <TM6>
F;256-332/Region: 11-residue repeats (P-D-N-T-D-D-N-G-P-Q-D)


  Query Match             10.7%;  Score 100.5;  DB 1;  Length 404;
  Best Local Similarity   23.6%;  Pred. No. 0.27;
  Matches   42;  Conservative   28;  Mismatches   53;  Indels   55;  Gaps    8;

Qy      16 LSFSFPLTMIALAISMSSWFNIWNNALSD-LGHAV----------------------KSSV 53
           || |  | :: | :::  |  |    :|| | |:                       :  :
Db      21 LSSSIGLALLLLLLALLFWLYI---VMSDWTGGALLVLYSFALMLIIIILIIFIFRRDLL 77

Qy      54 APIFNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYG-WIHFL-- 110
            |:  ||| :  | :::: | ||:  :   | ::   |. |         |:| ||:||
Db      78 CPLGGLGLLLLMITLLLIALWNLHGQALYLGIVLFIFGCLL--------VFGIWIYFLEI 129

Qy     111 -------------VSVLFFLSIIAYFIAISILDKSWIAV-----LLIIGHIAMW-YLH 149
                        : |||:||   ||: :    |  :      ||:   | :| | |
Db     130 LWRLGATLWQLLAFILAFFLAIILLIIALYLQQNWWTLLVDLLWLLLFMAILIWMYYH 187


RESULT  11
E82442
probable peptide ABC transporter, permease protein VCA0590 [imported] - Vibrio cholerae
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C;Accession: E82442
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers .
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: E82442
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <HEI>
A;Cross-references: GB:AE004389; GB:AE003853; NID:g9657989; PIDN:AAF96492.1; GSPDB:GN
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0590
A;Map position: 2
C;Superfamily: oligopeptide permease protein oppB


  Query Match             10.5%;  Score 98.5;  DB 2;  Length 360;
  Best Local Similarity   25.8%;  Pred. No. 0.36;
  Matches   34;  Conservative   29;  Mismatches   52;  Indels   17;  Gaps    6;

Qy       1 MSLNKHSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLG 60
           :|::   |  ::|:::  | ||  |: ||   | |::|::|:  :|:|:          |
Db     131 VSISLGLWSTLLIYLI--SIPLG-ISKAIHHGSRFDVWSSAVVIIGYAIP---------G 178

Qy      61 LAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSII 120
             |||::    | :||  ::| :  :          | |  | | |  :  |    :|
Db     179 FLFAIILIILFASGNYFSWFPLRGLVSDNFASLPWYQQVLD--YFW-HLTLPTL--AMVI 233

Qy     121 AYFIAISILDKS 132
             |  :|:| |:
Db     234 GGFATLSMLTKN 245


RESULT  12
A71694
hypothetical protein RP368 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: A71694
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499
A;Accession: A71694
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-280 <AND>
A;Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14827.1; PID:g386
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: RP368


  Query Match             10.4%;  Score 97.5;  DB 2;  Length 280;
  Best Local Similarity   27.7%;  Pred. No. 0.34;
  Matches   39;  Conservative   29;  Mismatches   50;  Indels   23;  Gaps    6;

Qy       3 LNKHSWMDMIIFILSFSFPLTMIALAISMSSW---FNIWNNALSDLGHAVKSSVAPIFN- 58
           |:  :|:  :| | ||   : :| |  ::  :   |    : | |||       | ::
Db     119 LDLDNWLVPLILINSFIIVIVLILLKGNLGLFLRAFGFNKDLLIDLG-----KPAELYRM 173

Qy      59 LGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSV----- 113
           |||:|   |  : |  :    ::| |   |:||  : |:|:   : |   || ::
Db     174 LGLSISNGLAALTGTLS----AQVNGFADINMGYGVALVGIGAIIIGRQIFLNNINNFNA 229

Qy     114 ---LF--FLSIIAYFIAISIL 129
              :|  |: |: |||::|||
Db     230 LKEIFACFIGILFYFISLSIL 250
```

```
RESULT  13
C82445
hypothetical protein VCA0550 [imported] - Vibrio cholerae (group O1 strain N16961)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C;Accession: C82445
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: C82445
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <HEI>
A;Cross-references: GB:AE004386; GB:AE003853; NID:g9657957; PIDN:AAF96452.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0550
A;Map position: 2


  Query Match             10.4%;  Score 97.5;  DB 2;  Length 360;
  Best Local Similarity   25.9%;  Pred. No. 0.43;
  Matches   45;  Conservative   29;  Mismatches   49;  Indels   51;  Gaps   10;

Qy       8 WMDMIIFILS--FSFPLTMIALAI----SMSSW-----FNIWNNALS-------DLGHAV 49
           | |::| |:   | | |::: : :         |     || : |  :        | || :
Db     197 WQDVVISIMGPCFGFLLSLLLVGVYWVTDSPFWAALAVFNAFLNLFNLLPVLPLDGGHIL 256

Qy      50 KSSVAPIFNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGV----FLNLIGVFDEVYG 105
              ||     |::   :| || |:  |  :         :|  |:|:    || |:|  | |:
Db     257 KS---VSFSMNSKMGVILCVLAILGGI--------ALSYSLGLTLFGFLLLMGALDIVFE 305

Qy     106 W-----IHFL--------VSVLFFLSIIAYFIAI-----SILDKSWIAVLLIIG 141
           |     | |         || ::: ::::  |||     |  |      |||:|
Db     306 WRQRHHSHLLPLNRYAQMVSTIWYFALVSGLIAIIIGFASTGDTLLSLPLLILG 359


RESULT  14
S26026
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 6 - Caenorhabditis elegans mitochondr
C;Species: mitochondrion Caenorhabditis elegans
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 20-Jun-2000
C;Accession: S26026; S25799
R;Okimoto, R.; Macfarlane, J.L.; Clary, D.O.; Wolstenholme, D.R.
Genetics 130, 471-498, 1992
A;Title: The mitochondrial genomes of two nematodes, Caenorhabditis elegans and Ascaris
A;Reference number: S26014; MUID:92201635
A;Accession: S26026
A;Molecule type: DNA
A;Residues: 1-144 <OKI>
A;Cross-references: EMBL:X54252; NID:g13988; PID:g2654243
R;Okimoto, R.; Macfarlane, J.L.; Wolstenholme, D.R.
Nucleic Acids Res. 18, 6113-6118, 1990
A;Title: Evidence for the frequent use of TTG as the translation initiation codon of mit
A;Reference number: S13139; MUID:91045077
A;Accession: S25799
A;Molecule type: DNA
A;Residues: 1-25 <OK2>
A;Cross-references: EMBL:X54252
C;Genetics:
A;Gene: ND6
A;Genome: mitochondrion
A;Genetic code: SGC4
A;Start codon: ATA
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 6
C;Keywords: mitochondrion; NAD; oxidoreductase


  Query Match             10.3%;  Score 97;  DB 2;  Length 144;
  Best Local Similarity   23.5%;  Pred. No. 0.21;
  Matches   31;  Conservative   23;  Mismatches   40;  Indels   38;  Gaps    4;

Qy      14 FILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGL 73
           | | ||   :|  :::||  ||:                 |   | : || :::|
Db      27 FFLIFSLLFSMPVISMSMHIWFS-----------------YFICLLFLSGIFVILVYF 67

Qy      74 RNLYSWSRVKGSL--------------IISMGVFLNLIGVFDEVYGWIHF---LVSVLF 115
             :|   : ||  :              :::   :| | | :  :| |  |   || :||
Db      68 SSLSKINVVKSYMAVFLLLLSMLYFSPTVLTYSSYLGLSGFYYSIY-WFIFCFILVCLLF 126

Qy     116 FLSIIAYFIAIS 127
           |::  :||:  |
Db     127 FMNFSSYFLNFS 138


RESULT  15
H82822
NADH-ubiquinone oxidoreductase, NQO14 subunit XF0318 [imported] - Xylella fastidiosa
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: H82822
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82822
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-485 <SIM>
A;Cross-references: GB:AE003884; GB:AE003849; NID:g9105127; PIDN:AAF83129.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0318
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 2


  Query Match             10.2%;  Score 96.5;  DB 2;  Length 485;
  Best Local Similarity   25.8%;  Pred. No. 0.68;
  Matches   41;  Conservative   22;  Mismatches   59;  Indels   37;  Gaps    6;

Qy      15 ILSFSFPLTMIALAISMSSWFNIWNNALS--DLGHAVKSSVAPIFNLGLAIGGILIVIVG 72
           :|  :| | || | |  :    :| : ||   ||  :   || :   | |  |:
Db      22 LLGSTFALLMIDLFIGQAR--KVWTHFLSVAILGVVLSMLVAGVGGQGAAFHGMF----- 74

Qy      73 LRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLF--FLSIIAYFIAISILD 130
                 |: |    | | : |:     |||| :     ||   : ::  |  : ::
Db      75 ---------VRDSAADVMKVVIVLVSALSLVYGWSYLRERNLFQGEIPVLVLFATLGMM- 124

Qy     131 KSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSF 169
                ||:  ||: | ||             ||||: |:
Db     125 -----VLVSAGHLLMVYLGL----------ELLALCSY 147


Search completed: May 30, 2001, 15:24:21
Job time: 91 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:24:05 ; Search time 9.55 Seconds
(without alignments)
660.001 Million cell updates/sec

Title: US-09-382-242-33
Perfect score: 942
Sequence: 1 MSLNKHSWMDMIIFILSFSF.....AVFSFLPFYIRDYFKSYTKR 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	100.5	10.7	404	1	LMP1_EBVC	P29362 epstein-bar
2	97.5	10.4	280	1	Y368_RICPR	Q9zdg1 rickettsia
3	97	10.3	144	1	NU6M_CAEEL	P24885 caenorhabdi
4	96	10.2	419	1	YHAP_BACSU	O07523 bacillus su
5	94.5	10.0	397	1	UT2_RAT	Q62668 rattus norv
6	94.5	10.0	1228	1	ALA5_ARATH	Q9sgg3 arabidopsis
7	94	10.0	394	1	CC91_YEAST	P41733 saccharomyc
8	93.5	9.9	460	1	MVIN_HELPJ	Q9zkw7 helicobacte
9	93	9.9	383	1	YJK8_YEAST	P42946 saccharomyc
10	90.5	9.6	532	1	YABM_BACSU	P37555 bacillus su
11	89.5	9.5	461	1	MVIN_HELPY	O25551 helicobacte
12	89	9.4	226	1	ATP6_APILI	Q00275 apis mellif
13	89	9.4	243	1	Y23B_MYCPN	P75314 mycoplasma
14	89	9.4	277	1	Y233_METJA	Q57685 methanococc
15	88.5	9.4	397	1	UT2_HUMAN	Q15849 homo sapien
16	88	9.3	313	1	NU1M_RHISA	O99824 rhipicephal
17	87.5	9.3	387	1	A2AB_MACPR	O19025 macroscelid
18	87.5	9.3	453	1	A2AB_RAT	P19328 rattus norv
19	87.5	9.3	455	1	A2AB_MOUSE	P30545 mus musculu
20	87.5	9.3	475	1	ARCD_LACSK	O53092 lactobacill
21	87.5	9.3	1202	1	ALAA_ARATH	Q91183 arabidopsis
22	87	9.2	261	1	COX3_GADMO	P55777 gadus morhu
23	87	9.2	261	1	FLIR_ERWCA	P34202 erwinia car
24	87	9.2	551	1	YIDE_HAEIN	P44472 haemophilus
25	86.5	9.2	198	1	COP_CLOPE	P18015 clostridium
26	86.5	9.2	305	1	YVBV_BACSU	O32256 bacillus su
27	86.5	9.2	364	1	YEJB_ECOLI	P33914 escherichia
28	86	9.1	275	1	TAGG_BACSU	P42953 bacillus su
29	86	9.1	299	1	Y440_METJA	Q57882 methanococc
30	86	9.1	319	1	Y189_MYCPN	P75262 mycoplasma
31	86	9.1	341	1	NU2M_DROYA	P03895 drosophila
32	86	9.1	369	1	NU1M_ACACA	Q37381 acanthamoeb
33	86	9.1	622	1	COX1_BACSU	P24010 bacillus su
34	85.5	9.1	492	1	NU4M_CHOCR	P48915 chondrus cr
35	85	9.0	279	1	DMSC_HAEIN	P45002 haemophilus
36	85	9.0	393	1	YHFC_ECOLI	P21229 escherichia
37	85	9.0	1029	1	YFC5_YEAST	P43571 saccharomyc
38	84.5	9.0	230	1	ATP6_ASTPE	Q33823 asterina pe
39	84.5	9.0	397	1	UT2_RABIT	Q28614 oryctolagus
40	84.5	9.0	430	1	RFBX_SALTY	P26400 salmonella
41	84.5	9.0	450	1	A2AB_HUMAN	P18089 homo sapien
42	84.5	9.0	474	1	ST24_SCHPO	Q10071 schizosacch
43	84.5	9.0	660	1	Y390_MYCPN	P75207 mycoplasma
44	84.5	9.0	1203	1	ALAB_ARATH	Q9saf5 arabidopsis
45	84	8.9	332	1	YPHD_ECOLI	P77315 escherichia

ALIGNMENTS

RESULT 1
LMP1_EBVC
ID LMP1_EBVC STANDARD; PRT; 404 AA.
AC P29362;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE LATENT MEMBRANE PROTEIN 1 (LMP-1) (P63).
GN BNLF1.
OS Epstein-barr virus (strain Cao) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=31525;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92013956; PubMed=1681026;
RA Hu L.F., Zabarovsky E.R., Chen F., Cao S.L., Ernberg I., Klein G.,
RA Winberg G.;
RT "Isolation and sequencing of the Epstein-Barr virus BNLF-1 gene
RT (LMP1) from a Chinese nasopharyngeal carcinoma.";
RL J. Gen. Virol. 72:2399-2409(1991).
CC -!- FUNCTION: THE LATENT MEMBRANE PROTEIN HAS TRANSFORMING ACTIVITY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X58140; CAA41148.1; -.
DR EMBL; D10059; BAA00948.1; -.
DR PIR; JQ1434; LABECA.
DR PIR; S21660; S21660.
KW Transmembrane; Phosphorylation; Transforming protein.
FT DOMAIN 1 24 CYTOPLASMIC.
FT TRANSMEM 25 44 POTENTIAL.
FT TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT DOMAIN 187 404 CYTOPLASMIC.
SQ SEQUENCE 404 AA; 43769 MW; D04536D3B65FF82E CRC64;

Query Match 10.7%; Score 100.5; DB 1; Length 404;
Best Local Similarity 23.6%; Pred. No. 1.1;
Matches 42; Conservative 28; Mismatches 53; Indels 55; Gaps 8;

```
Qy     16 LSFSFPLTMIALAISMSSWFNIWNNALSD-LGHAV----------------------KSSV 53
          || |  | :: | :::  |  |    :||  | |:                      :  :
Db     21 LSSSIGLALLLLLLALLFWLYI---VMSDWTGGALLVLYSFALMLIIIILIIFIFRRDLL 77

Qy     54 APIFNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYG-WIHFL-- 110
```

```
          |:  ||| :  | :::: | ||:  :   | ::   |  |        |:| ||:||
Db     78 CPLGGLGLLLLMITLLLIALWNLHGQALYLGIVLFIFGCLL--------VFGIWIYFLEI 129

Qy    111 ------------VSVLFFLSIIAYFIAISILDKSWIAV-----LLIIGHIAMW-YLH 149
                       : |||:||   ||: :    |  :     ||:   | :| | |
Db    130 LWRLGATLWQLLAFILAFFLAIILLIIALYLQQNWWTLLVDLLWLLLFMAILIWMYYH 187


RESULT   2
Y368_RICPR
ID   Y368_RICPR     STANDARD;      PRT;   280 AA.
AC   Q9ZDG1;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN RP368.
GN   RP368.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235271; CAA14827.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      3     23       POTENTIAL.
FT   TRANSMEM     52     72       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    123    143       POTENTIAL.
FT   TRANSMEM    196    216       POTENTIAL.
FT   TRANSMEM    233    253       POTENTIAL.
SQ   SEQUENCE   280 AA;  30554 MW;  CEF294794D843FA1 CRC64;


  Query Match             10.4%;  Score 97.5;  DB 1;  Length 280;
  Best Local Similarity   27.7%;  Pred. No. 1.3;
  Matches   39;  Conservative   29;  Mismatches   50;  Indels   23;  Gaps    6;

Qy      3 LNKHSWMDMIIFILSFSFPLTMIALAISMSSW---FNIWNNALSDLGHAVKSSVAPIFN- 58
          |:  :|:  :| | ||   : :| |  ::  :   |    : | |||       | ::
Db    119 LDLDNWLVPLILINSFIIVIVLILLKGNLGLFLRAFGFNKDLLIDLG-----KPAELYRM 173

Qy     59 LGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSV----- 113
          |||:|   |  : |  :    ::| |   |:||  : |:|:   : |  || ::
Db    174 LGLSISNGLAALTGTLS----AQVNGFADINMGYGVALVGIGAIIIGRQIFLNNINNFNA 229

Qy    114 ---LF--FLSIIAYFIAISIL 129
             :|  |: |: |||::|||
Db    230 LKEIFACFIGILFYFISLSIL 250


RESULT   3
NU6M_CAEEL
ID   NU6M_CAEEL     STANDARD;      PRT;   144 AA.
AC   P24885;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 6 (EC 1.6.5.3).
GN   ND6.
OS   Caenorhabditis elegans.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=92201635; PubMed=1551572;
RA   Okimoto R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;
RT   "The mitochondrial genomes of two nematodes, Caenorhabditis elegans
RT   and Ascaris suum.";
RL   Genetics 130:471-498(1992).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54252; CAA38153.1; -.
DR   PIR; S26026; S26026.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ   SEQUENCE   144 AA;  16788 MW;  04F51D0E354B40F9 CRC64;


  Query Match             10.3%;  Score 97;  DB 1;  Length 144;
  Best Local Similarity   23.5%;  Pred. No. 0.81;
  Matches   31;  Conservative   23;  Mismatches   40;  Indels   38;  Gaps    4;

Qy     14 FILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGL 73
          | | ||   :|  :::||  ||:                   |   | : || :::|
Db     27 FFLIFSLLFSMPVISMSMHIWFS-----------------YFICLLFLSGIFVILVYF 67

Qy     74 RNLYSWSRVKGSL--------------IISMGVFLNLIGVFDEVYGWIHF---LVSVLF 115
            :|   : ||  :              :::   :| | | :  :| | |   || :||
Db     68 SSLSKINVVKSYMAVFLLLLSMLYFSPTVLTYSSYLGLSGFYYSIY-WFIFCFILVCLLF 126

Qy    116 FLSIIAYFIAIS 127
          |::  :||:  |
Db    127 FMNFSSYFLNFS 138


RESULT   4
YHAP_BACSU
ID   YHAP_BACSU     STANDARD;      PRT;   419 AA.
AC   O07523;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 45.4 KDA PROTEIN IN SSPB-PRSA INTERGENIC REGION.
GN   YHAP.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: TO M.JANNASCHII MJ1024.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Y14078; CAA74423.1; -.
DR   EMBL; Z99109; CAB12830.1; -.
DR   SubtiList; BG12992; yhaP.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     16     36       POTENTIAL.
FT   TRANSMEM    186    206       POTENTIAL.
FT   TRANSMEM    235    255       POTENTIAL.
FT   TRANSMEM    283    303       POTENTIAL.
FT   TRANSMEM    318    338       POTENTIAL.
FT   TRANSMEM    340    360       POTENTIAL.
FT   TRANSMEM    369    389       POTENTIAL.
SQ   SEQUENCE   419 AA;  45429 MW;  E6434F27EE5E08A2 CRC64;


  Query Match             10.2%;  Score 96;  DB 1;  Length 419;
  Best Local Similarity   25.7%;  Pred. No. 2.3;
  Matches   46;  Conservative   35;  Mismatches   68;  Indels   30;  Gaps    9;

Qy      11 MIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFN-----LGLAIGG 65
           :::|:: ||  : | |  |:|     :     | :   : ||: ||       ||: : |
Db     189 IMLFVIYFS--VIMYASMIAM----EVATEKSSRVMEILISSMPPIQQMFAKLLGIGLVG 242

Qy      66 I--LIVIVGLRNLYSWSRVKGSLIISMGVFLNL--IGVFDEVYGWIHFLVSVLFFLSIIA 121
           |  | :|:|  :|      |     |:| ||||  :     :|  | ||::   : :: |
Db     243 ITQLAIIIGAGSLSLKLNQKSETASSVGAFLNLTDVSATTVIYAVIFFLLAYFSYATLAA 302

Qy     122 YF--IAISILDKSW----IAVLLIIGH-IAMWYLHFASEIPRGAAIPELLAVFSFLPFY 173
           :   :   | |       : :|:: |  |||: |:        |     : | ||:||:
Db     303 FLGSVVSRIEDVQQTITPMTLLVVAGFMIAMFGLN--------APDAGFITVTSFIPFF 353


RESULT   5
UT2_RAT
ID   UT2_RAT        STANDARD;      PRT;   397 AA.
AC   Q62668;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UREA TRANSPORTER, KIDNEY.
GN   SLC14A2 OR UT2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;
RX   MEDLINE=95386714; PubMed=7657826;
RA   Smith C.P., Lee W.-S., Martial S., Knepper M.A., You G., Sands J.M.,
RA   Hediger M.A.;
RT   "Cloning and regulation of expression of the rat kidney urea
RT   transporter (rUT2).";
RL   J. Clin. Invest. 96:1556-1563(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney inner medulla;
RX   MEDLINE=97117152; PubMed=8958221;
RA   Shayakul C., Steel A., Hediger M.A.;
RT   "Molecular cloning and characterization of the vasopressin-regulated
RT   urea transporter of rat kidney collecting ducts.";
RL   J. Clin. Invest. 98:2580-2587(1996).
CC   -!- FUNCTION: SPECIALIZED LOW-AFFINITY UREA TRANSPORTER. MEDIATES UREA
CC       TRANSPORT IN KIDNEY.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: KIDNEY.
CC   -!- SIMILARITY: TO OTHER MAMMALIAN UREA TRANSPORTERS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U09957; AAA84392.1; -.
DR   EMBL; U77971; AAB50197.1; ALT_INIT.
KW   Transport; Transmembrane; Glycoprotein.
FT   TRANSMEM     68     85       POTENTIAL.
FT   TRANSMEM     92    109       POTENTIAL.
FT   TRANSMEM    115    135       POTENTIAL.
FT   TRANSMEM    143    163       POTENTIAL.
FT   TRANSMEM    172    192       POTENTIAL.
FT   TRANSMEM    239    257       POTENTIAL.
FT   TRANSMEM    264    280       POTENTIAL.
FT   TRANSMEM    287    303       POTENTIAL.
FT   TRANSMEM    309    329       POTENTIAL.
FT   TRANSMEM    331    351       POTENTIAL.
FT   CARBOHYD    210    210       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   397 AA;  43237 MW;  EA2068EA25661224 CRC64;


  Query Match             10.0%;  Score 94.5;  DB 1;  Length 397;
  Best Local Similarity   22.2%;  Pred. No. 2.8;
  Matches   50;  Conservative   30;  Mismatches   60;  Indels   85;  Gaps   12;

Qy       4 NKHSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAI 63
           | : |: : : ::| : |:   ||:   | |          ||          |:| | |
Db     140 NYYWWLLLPVIVMSMTCPILSSALSTVFSKW---------DL---------PVFTLPFNI 181

Qy      64 GGIL-IVIVGLRNLY---------------SWSRVKGSLI---ISMGVFLNLIGVFDEVY 104
             | :   |  ||:                :|| |:  |:   | :|:          :||
Db     182 AVTLYLAATGHYNLFFPTKLLQPAVTTPNITWSDVQVPLLLRAIPVGI--------GQVY 233

Qy     105 G----WIHFLVSVLFFL------------SIIAYFIAISILD------------KSWIAV 136
           |    |   :  |  |:             | |    |:||               | :|
Db     234 GCDNPWTGGIFLVALFVSSPLICLHAAIGSTIGMLAALSIATPFDSIYFGLCGFNSTLAC 293

Qy     137 LLIIG--HIAMWYLH--------FASEIPRGAAIPELLAVFSFLP 171
           : | |  ::  |  |         ||: :  |||:  :|:||   |
Db     294 IAIGGMFYVITWQTHLLAIACALFAAYL--GAALANMLSVFGLPP 336


RESULT   6
ALA5_ARATH
ID   ALA5_ARATH     STANDARD;      PRT;  1228 AA.
AC   Q9SGG3;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE 5 (EC 3.6.3.13).
GN   ALA5 OR AT1G72700 OR F28P22.11.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
```

```
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- FUNCTION: INVOLVED IN TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC       (E1-E2 ATPASES). SUBFAMILY IV.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC010926; AAF21158.1; -.
DR   InterPro; IPR001757; -.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Metal-binding;
KW   Magnesium; Multigene family.
FT   DOMAIN        1     74       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     75     96       POTENTIAL.
FT   DOMAIN       97    100       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    101    123       POTENTIAL.
FT   DOMAIN      124    305       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    306    327       POTENTIAL.
FT   DOMAIN      328    359       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    360    377       POTENTIAL.
FT   DOMAIN      378    934       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    935    954       POTENTIAL.
FT   DOMAIN      955    968       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    969    988       POTENTIAL.
FT   DOMAIN      989   1018       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1019   1040       POTENTIAL.
FT   DOMAIN     1048   1054       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1055   1077       POTENTIAL.
FT   DOMAIN     1078   1083       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1084   1104       POTENTIAL.
FT   DOMAIN     1105   1117       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1118   1146       POTENTIAL.
FT   DOMAIN     1147   1228       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     425    425       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL       879    879       MAGNESIUM (BY SIMILARITY).
FT   METAL       883    883       MAGNESIUM (BY SIMILARITY).
SQ   SEQUENCE   1228 AA;  139340 MW;  47E47A9297131CB8 CRC64;


  Query Match           10.0%;  Score 94.5;  DB 1;  Length 1228;
  Best Local Similarity  25.4%;  Pred. No. 7;
  Matches   47;  Conservative   33;  Mismatches   60;  Indels   45;  Gaps   10;

Qy      14 FILSFSFPLTMIAL-AISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIV- 71
           |  : :| ||:    | :  |  :::|:    | : | :|: |:  ||:    :   |
Db     941 FYKNIAFGLTLFYFEAFTGFSGQSVYNDYYLLLFNVVLTSL-PVIALGVFEQDVSSEICL 999

Qy      72 --------GLRNL-YSWSRVKG--------SLIISMGVFLNLIGVFDEVY---GWIHFL- 110
                   | :|| : |||: |        ||:|    |||:  :: : :   |    :
Db    1000 QFPALYQQGTKNLFFDWSRILGWMCNGVYASLVI---FFLNIGIIYSQAFRDNGQTADMD 1056

Qy     111 -VSVLFFLSII---AYFIAISILDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAV 166
            |    |  ||      ||:::   :|| :|| | | ||||               :|:
Db    1057 AVGTTMFTCIIWAANVQIALTMSHFTWIQHVLIWGSIGMWYL--------------FVAI 1102

Qy     167 FSFLP 171
           :| :|
Db    1103 YSMMP 1107


RESULT   7
CC91_YEAST
ID   CC91_YEAST     STANDARD;      PRT;   394 AA.
AC   P41733;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CELL DIVISION CONTROL PROTEIN 91.
GN   CDC91 OR YLR459W OR L9122.2.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bi E., Pringle J.R.;
RL   Submitted (XXX-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31649; AAA34487.1; -.
DR   EMBL; U22383; AAB64722.1; -.
DR   SGD; S0004451; CDC91.
KW   Cell division; Cell cycle.
SQ   SEQUENCE   394 AA;  44741 MW;  75F4BD39FF871A58 CRC64;


  Query Match           10.0%;  Score 94;  DB 1;  Length 394;
  Best Local Similarity  22.7%;  Pred. No. 3;
  Matches   48;  Conservative   40;  Mismatches   65;  Indels   58;  Gaps   11;

Qy       8 WMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDL-------GHAVKSSVAPIFNLG 60
           |:  ::: ::   |||:::  || ||     | |:|         |: : |||    :
Db     114 WLPGLLYAVN---PLTLLS-CISRSS-IIFTNFAISSSLYCILAEGNVLLSSVMISISGY 168

Qy      61 LAIGGILIVIVGLRNLYSW-SRVKGSLIISMGVFLNLI-----------GVFDEVYG--- 105
           |::  ||::|  |  | ||  |:  :::  : : : |:              :|||
Db     169 LSVYPILLLIPLLGMLKSWRQRILSAIVSILSLLILLLFSYSILGSQSWSFLTQVYGSII 228

Qy     106 ----------WIHFLVSVL-----FFLSIIAYFIAISIL-------DKSWIAVLLIIGH 142
                     | :| : :      || ::   |||: |         : : | :| ||
Db     229 TFEKVFPNLGLWWYFFIEMFDTFIPFFKAVFNIFIAVFITPFTLRYHKQPFYAFILCIGW 288

Qy     143 IAMWYLHFASEIPRGAAIPELLAVFSFLPFY 173
           | :         |   :: :    ||||||:
Db     289 IVL-----TKPYP---SLGDAGFFFSFLPFF 311


RESULT   8
MVIN_HELPJ
ID   MVIN_HELPJ     STANDARD;      PRT;   460 AA.
```

```
AC   Q9ZKW7;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   VIRULENCE FACTOR MVIN HOMOLOG.
GN   MVIN OR JHP0817.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MVIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001511; AAD06398.1; -.
KW   Transmembrane.
FT   TRANSMEM      4     24       POTENTIAL.
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM     95    115       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    155    175       POTENTIAL.
FT   TRANSMEM    228    248       POTENTIAL.
FT   TRANSMEM    258    278       POTENTIAL.
FT   TRANSMEM    292    312       POTENTIAL.
FT   TRANSMEM    336    356       POTENTIAL.
FT   TRANSMEM    366    386       POTENTIAL.
FT   TRANSMEM    387    407       POTENTIAL.
FT   TRANSMEM    428    448       POTENTIAL.
SQ   SEQUENCE   460 AA;  51484 MW;  0E957BB6023C1FC1 CRC64;


  Query Match             9.9%;  Score 93.5;  DB 1;  Length 460;
  Best Local Similarity   22.5%;  Pred. No. 3.7;
  Matches   31;  Conservative   33;  Mismatches   55;  Indels   19;  Gaps    6;

Qy      15 ILSFSFPLTMIA-LAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGL 73
           | |  | | ::  |   ::  |::|  |  :   | | |:  :| ||||    |: ::|:
Db     332 ITSQVFSLYLLGLLPFGLTKLFSLWLYAKLEQKKAAKISLISLF-LGLAASLSLMPLLGV 390

Qy      74 RNLYSWSRVKGSLIISMGV----FLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISIL 129
             |   : : |  :  : :    | : :|:   :  |:     |: ||: :   : ::
Db     391 LGLALANSLSGLFLFVLTIKAFGFQSFLGIIKNLKSWL-----VILFLACVEILLLLAF- 444

Qy     130 DKSWIAVLLIIGHIAMWY 147
           |||:       |: ::|
Db     445 -KSWVT------HLYLFY 455


RESULT   9
YJK8_YEAST
ID   YJK8_YEAST     STANDARD;      PRT;   383 AA.
AC   P42946;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 41.5 KDA PROTEIN IN GZF3-IME2 INTERGENIC REGION.
GN   YJL108C OR J0811.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=96090136; PubMed=7483851;
RA   Rasmussen S.W.;
RT   "A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1
RT   and CSD3 genes, a TCP-1-related gene, an open reading frame similar
RT   to the DAL80 gene, and a tRNA(Arg).";
RL   Yeast 11:873-883(1995).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: TO S.POMBE SPAC16A10.01.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85021; CAA59386.1; -.
DR   EMBL; Z49383; CAA89402.1; -.
DR   SGD; S0003644; YJL108C.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     66     86       POTENTIAL.
FT   TRANSMEM     88    108       POTENTIAL.
FT   TRANSMEM    118    138       POTENTIAL.
FT   TRANSMEM    142    162       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
FT   TRANSMEM    217    237       POTENTIAL.
FT   TRANSMEM    242    262       POTENTIAL.
FT   TRANSMEM    272    292       POTENTIAL.
FT   TRANSMEM    294    314       POTENTIAL.
FT   TRANSMEM    353    373       POTENTIAL.
SQ   SEQUENCE   383 AA;  41488 MW;  2030F73D1823CE04 CRC64;


  Query Match             9.9%;  Score 93;  DB 1;  Length 383;
  Best Local Similarity   21.6%;  Pred. No. 3.4;
  Matches   51;  Conservative   40;  Mismatches   59;  Indels   86;  Gaps   12;

Qy       8 WMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGI- 66
           |::: |   ||   | : :|   :|    :::|       : :|:   |  | | | |
Db      88 WVNLAI---SFFMGLCVGSLQFILSQKSYMYSNVFE-----ISASIVVSF-CGRAFGSIP 138

Qy      67 -------------------LIVIVGLRNLYSWSRVKG----------SLIISMGVFLNL 96
                              |:: |   | | | | |          || :  |: |
Db     139 RSHICFGAVTQGSLALILPGYIILCGALELQSRSLVAGAVRMFYAIIYSLFLGFGITLG- 197

Qy      97 IGVFDEVYGWIHF----------LVSVLF-FLSIIAYFIAISILDKSWIAVLLII----- 140
                ::||::           |:|  | || : |: |:||:|::: |: | ::
Db     198 ----SALFGWMYHNATNEISCPQLISPWFRFLFVPAFTISISLLNQAHISQLPVMVFISC 253

Qy     141 -GHIAMWYL--HFA--------------------SEIPRGAAIPELL-AVFSFLP 171
            |::  ::   |||                    | | :| |:  :| |:|  :|
Db     254 TGYVVTYWAGKHFANSTEFTAALAAFVIGVLGNLYSRIWKGLAVSAMLPAIFVQVP 309


RESULT  10
YABM_BACSU
ID   YABM_BACSU     STANDARD;      PRT;   532 AA.
AC   P37555;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   HYPOTHETICAL 57.4 KDA PROTEIN IN MFD-DIVIC INTERGENIC REGION.
GN   YABM.
```

```
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96051385; PubMed=7584024;
RA   Ogasawara N., Nakai S., Yoshikawa H.;
RT   "Systematic sequencing of the 180 kilobase region of the Bacillus
RT   subtilis chromosome containing the replication origin.";
RL   DNA Res. 1:1-14(1994).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: STRONG, TO B.SUBTILIS SPOVB.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26185; BAA05292.1; -.
DR   EMBL; Z99104; CAB11833.1; -.
DR   SubtiList; BG10120; yabM.
DR   InterPro; IPR002797; -.
DR   Pfam; PF01943; Polysacc_synt; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     25     45       POTENTIAL.
FT   TRANSMEM     65     85       POTENTIAL.
FT   TRANSMEM    109    129       POTENTIAL.
FT   TRANSMEM    134    154       POTENTIAL.
FT   TRANSMEM    179    199       POTENTIAL.
FT   TRANSMEM    203    223       POTENTIAL.
FT   TRANSMEM    248    268       POTENTIAL.
FT   TRANSMEM    302    322       POTENTIAL.
FT   TRANSMEM    344    364       POTENTIAL.
FT   TRANSMEM    371    391       POTENTIAL.
FT   TRANSMEM    392    412       POTENTIAL.
FT   TRANSMEM    425    445       POTENTIAL.
FT   TRANSMEM    459    479       POTENTIAL.
FT   TRANSMEM    494    514       POTENTIAL.
SQ   SEQUENCE   532 AA;  57357 MW;  DAAACAF32903E553 CRC64;


  Query Match             9.6%;  Score 90.5;  DB 1;  Length 532;
  Best Local Similarity   24.7%;  Pred. No. 6.7;
  Matches   53;  Conservative   35;  Mismatches   84;  Indels   43;  Gaps   10;

Qy       5 KHSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAI- 63
           |  |:    |:|  :  :| |  |:    : ||    : |:|  :   | |   ||:|:
Db       8 KRHWIWQGAFVLILAGVITKILSAVYRVPFQNI----VGDVGFYIYQQVYPF--LGIAVM 61

Qy      64 ---GGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGV--FDEVY----------GWIH 108
              |  ::|  | | ||      | ||  :||:|||:  |  :|           |  |
Db      62 LSTSGFPVIISKLMNDYSEKNHHTILKIS-ALFLSLIGILLFLCLYLGAVPIALFMGDSH 120

Qy     109 F--LVSVLFFLSIIAYFI---------------AISILDKSWIAVLLIIGHIAMWYL-H 149
             |: |  :  ::  |:                 |:| : : :: | :::| :: | :
Db     121 LAVLIQVAAYAFLLFPFVALLRGGFQGRHEMLPSALSQMTEQFLRVAVLLG-LSFWLVKK 179

Qy     150 FASEIPRGAAIPELLAVFSFLPFYIRDYFKSYTKR 184
           ||    |||        | :   |  :|   |||
Db     180 GASLYTAGAAAASGSLAGSLVALIILGFFWFKTKR 214


RESULT  11
MVIN_HELPY
ID   MVIN_HELPY     STANDARD;      PRT;   461 AA.
AC   O25551;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   VIRULENCE FACTOR MVIN HOMOLOG.
GN   MVIN OR HP0885.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MVIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000598; AAD07933.1; -.
DR   TIGR; HP0885; -.
KW   Transmembrane.
FT   TRANSMEM      5     25       POTENTIAL.
FT   TRANSMEM     51     71       POTENTIAL.
FT   TRANSMEM     96    116       POTENTIAL.
FT   TRANSMEM    123    143       POTENTIAL.
FT   TRANSMEM    156    176       POTENTIAL.
FT   TRANSMEM    229    249       POTENTIAL.
FT   TRANSMEM    259    279       POTENTIAL.
FT   TRANSMEM    294    314       POTENTIAL.
FT   TRANSMEM    337    357       POTENTIAL.
FT   TRANSMEM    372    392       POTENTIAL.
FT   TRANSMEM    402    422       POTENTIAL.
FT   TRANSMEM    429    449       POTENTIAL.
SQ   SEQUENCE   461 AA;  51618 MW;  62B7DA434967A46C CRC64;


  Query Match             9.5%;  Score 89.5;  DB 1;  Length 461;
  Best Local Similarity   22.5%;  Pred. No. 7.1;
  Matches   31;  Conservative   32;  Mismatches   56;  Indels   19;  Gaps    6;

Qy      15 ILSFSFPLTMIA-LAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGL 73
           | |  | | ::  |   ::  |::|  |  :   | | |:  :| ||||    |: ::|:
Db     333 ITSQVFSLYLLGLLPFGLTKLFSLWLYAKLEQKKAAKISLISLF-LGLAASLSLMPLLGV 391

Qy      74 RNLYSWSRVKGSLIISMGV----FLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISIL 129
             |   : : |  :  : :     |   :|:   :  |:     |: ||: :   : ::
Db     392 LGLALANSLSGLFLFVLTIKAFGFQLFLGIIKNLKSWL-----VILFLACVEILLLLAF- 445

Qy     130 DKSWIAVLLIIGHIAMWY 147
           |||:       |: ::|
Db     446 -KSWVT------HLYLFY 456


RESULT  12
ATP6_APILI
ID   ATP6_APILI     STANDARD;      PRT;   226 AA.
AC   Q00275;
```

```
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
GN   ATP6.
OS   Apis mellifera ligustica (Common honeybee).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC   Apoidea; Apidae; Apis.
OX   NCBI_TaxID=7469;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Thorax;
RX   MEDLINE=92261310; PubMed=1533894;
RA   Crozier R.H., Crozier Y.C.;
RT   "The cytochrome b and ATPase genes of honeybee mitochondrial DNA.";
RL   Mol. Biol. Evol. 9:474-482(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Thorax;
RX   MEDLINE=93114603; PubMed=8417993;
RA   Crozier R.H., Crozier Y.C.;
RT   "The mitochondrial genome of the honeybee Apis mellifera: complete
RT   sequence and genome organization.";
RL   Genetics 133:97-117(1993).
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L06178; AAB96802.1; -.
DR   EMBL; M87065; AAA31634.1; -.
DR   InterPro; IPR000568; -.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
SQ   SEQUENCE   226 AA;  26694 MW;  B6F78A57B06A1FD4 CRC64;


  Query Match             9.4%;  Score 89;  DB 1;  Length 226;
  Best Local Similarity   24.0%;  Pred. No. 4.3;
  Matches   37;  Conservative   42;  Mismatches   51;  Indels   24;  Gaps    9;

Qy       1 MSLNKHSWMDMIIFILSFSFPLT---MIALAISMSSWFN-----IWNNALSDLGHAVKSS 52
           :||  :  :  |  :: : | ||   :: : :|:: ||:     |:|| :  | | |  :
Db      77 ISLMLYIMITNIFSLIPYVFTLTSHLLLNMILSLTLWFSFLIYLIYNNYIMFLSHLVPLN 136

Qy      53 VAPIFNLGLAIGGILIVIVGLRNLY--SWS---RVKGSLIISMGVFLNLIGVFDEVYGWI 107
             :|:| :        :||: | :|    |:   |: :| ||  : | |:|:|  :  :|
Db     137 -SPVFLMN------FMVIIELISLIIRPWTLSIRLSANL-ISGHLILTLLGIF--ISNFI 186

Qy     108 HFL-VSVLFFLSIIAYFIAISILDKSWIAVLLI 140
             | ::::    ::   | :|::     ::|||:
Db     187 SILPINLMIQNMLLTLEIFMSMIQSYVFSILLIL 220


RESULT  13
Y23B_MYCPN
ID   Y23B_MYCPN     STANDARD;      PRT;   243 AA.
AC   P75314;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN MG323.1 HOMOLOG (P01_ORF243).
GN   MPN469 OR MP372.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000035; AAB96020.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     38     58       POTENTIAL.
FT   TRANSMEM     99    119       POTENTIAL.
FT   TRANSMEM    143    163       POTENTIAL.
FT   TRANSMEM    204    224       POTENTIAL.
SQ   SEQUENCE   243 AA;  28278 MW;  DFD7355C3DD876FE CRC64;


  Query Match             9.4%;  Score 89;  DB 1;  Length 243;
  Best Local Similarity   24.7%;  Pred. No. 4.5;
  Matches   36;  Conservative   23;  Mismatches   41;  Indels   46;  Gaps    8;

Qy      13 IFILSFSFPLTMIALAI----------SMSSWFNIWNNALSDLGHAVKSSVAPIFNL 59
           :|:|||     |  : |             :| ::  ||         ||   | |||
Db      41 LFLLSFFVTAVMFLVGIFHTEQFTLNDQNQGISGFYLFWN---------VKKP-ADIFNA 90

Qy      60 GLAIG----GILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVF--DEVYGWIHFLVSV 113
                 || |: :|| :|:       |:| :| :   | :|  .::  |   : |
Db      91 NFVYSISSFGIAILALGLFSLF--------LMIFLG-YRWAISLFIKSQITKWERVIFST 141

Qy     114 LFFLSIIAYFIAISILDKSWIAVLLI 139
            |: |:::||          |||::|:
Db     142 GFYFSVVAYCF--------WIALMLL 159


RESULT  14
Y233_METJA
ID   Y233_METJA     STANDARD;      PRT;   277 AA.
AC   Q57685;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0233.
GN   MJ0233.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
```

```
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: TO M.JANNASCHII MJ1189.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67479; AAB98226.1; -.
DR   TIGR; MJ0233; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     23     43       POTENTIAL.
FT   TRANSMEM     61     81       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    144    164       POTENTIAL.
FT   TRANSMEM    197    217       POTENTIAL.
FT   TRANSMEM    221    241       POTENTIAL.
FT   TRANSMEM    243    263       POTENTIAL.
SQ   SEQUENCE   277 AA;  31306 MW;  31F66D38BD552B39 CRC64;
```

```
  Query Match             9.4%;  Score 89;  DB 1;  Length 277;
  Best Local Similarity   22.3%;  Pred. No. 5.1;
  Matches   43;  Conservative   27;  Mismatches   53;  Indels   70;  Gaps    8;

Qy      44 DLGHAVKSSV---APIFNLGLAIGGIL---------IVIVGLRNLYSWSRV-----KGSL 86
           |: | : | :     || :|| :  |:          : ||      |: :     :| |
Db      53 DIMHIIMSFLIYFGIIFIIGLIVSAIIGGYNVRIMKTTVEGLNVAPDWNNITDLLYRGIL 112

Qy      87 IISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSII---------AYFIAISILDKSWIAVL 137
            |   | ||:|          :|: ::||   |          |: | ||||       :
Db     113 YIVGLVLLNII---------FYFIPAILFVFGIFSLYISKIIGAFLIIISIL-------I 156

Qy     138 LIIGHIAMWYLHFASEIP--------------------RGAAIPELLAVFSF------- 169
            ||  |::|     :|:                     |   :  ::|: :|
Db     157 FIISVISLWLYSKLAEVNYSVKGFYGFFEFKEIFKMIGIRYIILVIIIAIINFIISLIVV 216

Qy     170 LPFYIRDYFKSYT 182
           ||  | | | ||:
Db     217 LPLNIIDIFISYS 229
```

```
RESULT  15
UT2_HUMAN
ID   UT2_HUMAN      STANDARD;      PRT;   397 AA.
AC   Q15849;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UREA TRANSPORTER, KIDNEY.
GN   SLC14A2 OR UT2 OR HUT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=96228053; PubMed=8647271;
RA   Olives B., Sonia M., Mattei M.-G., Matassi G., Rousselet G.,
RA   Ripoche P., Cartron J.-P., Bailly P.;
RT   "Molecular characterization of a new urea transporter in the human
RT   kidney.";
RL   FEBS Lett. 386:156-160(1996).
CC   -!- FUNCTION: SPECIALIZED LOW-AFFINITY UREA TRANSPORTER. MEDIATES UREA
CC       TRANSPORT IN KIDNEY.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: KIDNEY.
CC   -!- SIMILARITY: TO OTHER MAMMALIAN UREA TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X96969; CAA65657.1; -.
DR   MIM; 601611; -.
KW   Transport; Transmembrane; Glycoprotein.
FT   TRANSMEM     68     85       POTENTIAL.
FT   TRANSMEM     92    109       POTENTIAL.
FT   TRANSMEM    115    135       POTENTIAL.
FT   TRANSMEM    143    163       POTENTIAL.
FT   TRANSMEM    172    192       POTENTIAL.
FT   TRANSMEM    239    257       POTENTIAL.
FT   TRANSMEM    264    280       POTENTIAL.
FT   TRANSMEM    287    303       POTENTIAL.
FT   TRANSMEM    309    329       POTENTIAL.
FT   TRANSMEM    331    351       POTENTIAL.
FT   CARBOHYD    210    210       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   397 AA;  43395 MW;  310AD64B1DFA9623 CRC64;
```

```
  Query Match             9.4%;  Score 88.5;  DB 1;  Length 397;
  Best Local Similarity   21.7%;  Pred. No. 7.3;
  Matches   48;  Conservative   29;  Mismatches   59;  Indels   85;  Gaps   12;

Qy       8 WMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGIL 67
           |: : : |:| | |:   ||    | |         ||         |:| |   |  | |
Db     144 WLLLPVIIMSMSCPILSSALGTIFSKW---------DL---------PVFTLPFNITVTL 185

Qy      68 -IVIVGLRNLY--------------SWSRVKGSLI---ISMGVFLNLIGVFDEVYG--- 105
            :   | ||:                :|| |:  |:   | :|:         :|||
Db     186 YLAATGHYNLFFPTTLLQPASAMPNITWSEVQVPLLLRAIPVGI--------GQVYGCDN 237

Qy     106 -WIHFLVSVLFFL------------SIIAYFIAISILD------------KSWIAVLLII 140
            |   :  |:          | :    |::|                  | :| : |
Db     238 PWTGGIFLIALFISSPLICLHAAIGSTMGMLAALTIATPFDSIYFGLCGFNSTLACIAIG 297

Qy     141 G--HIAMWYLH--------FASEIPRGAAIPELLAVFSFLP 171
           |  ::  |  |         ||: :  |||:  :|:||   |
Db     298 GMFYVITWQTHLLAIACALFAAYL--GAALANMLSVFGLPP 336
```

```
Search completed: May 30, 2001, 15:26:02
Job time: 117 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:23:50 ; Search time 22.38 Seconds
(without alignments)
963.640 Million cell updates/sec

Title: US-09-382-242-33
Perfect score: 942
Sequence: 1 MSLNKHSWMDMIIFILSFSF.....AVFSFLPFYIRDYFKSYTKR 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	163.5	17.4	175	1	Q9V235	Q9v235 pyrococcus
2	144.5	15.3	156	1	O59449	O59449 pyrococcus
3	142	15.1	192	1	O58000	O58000 pyrococcus
4	129.5	13.7	183	1	O59239	O59239 pyrococcus
5	126.5	13.4	156	1	Q9V1S7	Q9v1s7 pyrococcus
6	118	12.5	165	1	Q9Y8U9	Q9y8u9 aeropyrum p
7	109.5	11.6	512	2	Q9RQP4	Q9rqp4 burkholderi
8	105.5	11.2	401	2	Q9KS98	Q9ks98 vibrio chol
9	104.5	11.1	373	2	Q9RXQ6	Q9rxq6 deinococcus
10	103.5	11.0	600	2	Q9KNN1	Q9knn1 vibrio chol
11	102.5	10.9	293	5	Q9VGY0	Q9vgy0 drosophila
12	101.5	10.8	365	8	Q9XMT1	Q9xmt1 tetrahymena
13	98.5	10.5	360	2	Q9KLZ9	Q9klz9 vibrio chol
14	97.5	10.4	231	8	Q9ZZU5	Q9zzu5 littorina s
15	97.5	10.4	360	2	Q9KM39	Q9km39 vibrio chol
16	97.5	10.4	565	8	Q9MR91	Q9mr91 loligo blee
17	97.5	10.4	566	8	Q9MJ59	Q9mj59 loligo blee
18	96.5	10.2	485	2	Q9PGI2	Q9pgi2 xylella fas
19	95.5	10.1	371	14	Q9YLF8	Q9ylf8 human herpe
20	95	10.1	433	2	Q9S4G6	Q9s4g6 leptospira
21	95	10.1	921	10	Q9SGH9	Q9sgh9 arabidopsis
22	95	10.1	1320	5	Q27930	Q27930 drosophila
23	94.5	10.0	348	8	Q9TD57	Q9td57 cynopoecilu
24	94.5	10.0	1228	10	Q9SGG3	Q9sgg3 arabidopsis
25	94	10.0	510	2	Q9ZDE9	Q9zde9 rickettsia
26	93.5	9.9	393	2	Q9L4R8	Q914r8 bacillus ce
27	93.5	9.9	587	5	Q9VGP3	Q9vgp3 drosophila
28	93.5	9.9	870	5	Q9XTQ4	Q9xtq4 caenorhabdi
29	93	9.9	487	2	O67873	O67873 aquifex aeo
30	92.5	9.8	303	2	O34706	O34706 bacillus su
31	92.5	9.8	395	2	Q44575	Q44575 acetobacter
32	92	9.8	407	2	O66504	O66504 aquifex aeo
33	91.5	9.7	267	2	Q9RNL1	Q9rnl1 zymomonas m
34	91.5	9.7	419	10	Q9M0Y1	Q9m0y1 arabidopsis
35	91.5	9.7	452	2	Q9ZD69	Q9zd69 rickettsia
36	91	9.7	273	2	O66090	O66090 lactococcus
37	91	9.7	313	8	O99816	O99816 ixodes hexa
38	91	9.7	425	8	Q9MD45	Q9md45 schistosoma
39	91	9.7	435	2	Q9ZK40	Q9zk40 helicobacte
40	91	9.7	563	2	Q56876	Q56876 yersinia en
41	91	9.7	643	1	O29273	O29273 archaeoglob
42	90.5	9.6	442	2	O25239	O25239 helicobacte
43	90	9.6	274	8	Q36360	Q36360 drosophila
44	90	9.6	381	14	Q89558	Q89558 human herpe
45	90	9.6	517	8	Q9TAK5	Q9tak5 cafeteria r

ALIGNMENTS

RESULT 1
Q9V235
ID Q9V235 PRELIMINARY; PRT; 175 AA.
AC Q9V235;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 19.1 KDA PROTEIN.
GN PAB0162.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ248283; CAB49163.1; -.
KW Hypothetical protein.
SQ SEQUENCE 175 AA; 19133 MW; 19AB460C3DC15F79 CRC64;

Query Match 17.4%; Score 163.5; DB 1; Length 175;
Best Local Similarity 34.8%; Pred. No. 0.00013;
Matches 55; Conservative 28; Mismatches 64; Indels 11; Gaps 5;

```
Qy      24 MIALA------ISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRNLY 77
           :||||      :    |::|  ||:|||| |      : | || |||:|::  |   |
Db      10 IIALAGVVFAYLLNRDWWSITENAISDLGR-VGLPNNWVMNCGLIIGGLLLLGYGAWRLK 68

Qy      78 SWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSIIA--YFIAISILDKSWIA 135
              :: :   ||| ||||| |   | |: ||  ||:|: |    |:|||| || :
Db      69 KSKNPGWAIYVVGSVFLVLIGVFPEGTEW-HYEVSWGFFVSMFAAMLIISISILKKSRLG 127

Qy     136 VL-LIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPF 172
            | | :  :::    |: :   | |: | ::: :|| |
Db     128 FLGLAVFCVSLPLALFSLKAFSGVAVAETVSILAFLIF 165
```

RESULT 2

```
O59449
ID   O59449      PRELIMINARY;      PRT;   156 AA.
AC   O59449;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 17.5 KDA PROTEIN PH1785.
GN   PH1785.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000007; BAA30904.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   156 AA;  17521 MW;  898EACD3F85401F6 CRC64;


  Query Match             15.3%;  Score 144.5;  DB 1;  Length 156;
  Best Local Similarity   31.3%;  Pred. No. 0.0025;
  Matches   52;  Conservative   30;  Mismatches   59;  Indels   25;  Gaps    8;

Qy      12 IIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNL---GLAIGGILI 68
           |: :: |   :  :|| || : ||:   |||||:| | |:    :|||   || | ||::
Db       7 IVMVIYF---IIGLALVISQNPWFSFTKNALSDMGSA-KNPKGWMFNLYIIGLGIIGIIV 62

Qy      69 VIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISI 128
             :  |||         | ||| : | |:|:| |     |   ::| :|     || : :
Db      63 ANLLNRNL---------LKISM-ILLILVGIFPEEEP-PHTPSAILMYL---LSFIDMGL 108

Qy     129 LDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYI 174
           ::|  :  |   | :  : |      | |||||:   : | : :
Db     109 YSRTWRIIATITFTIMVILVSFKI----GLAIPELIGAAAILSYIL 150


RESULT   3
O58000
ID   O58000      PRELIMINARY;      PRT;   192 AA.
AC   O58000;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 21.5 KDA PROTEIN PH0262.
GN   PH0262.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000001; BAA29334.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   192 AA;  21473 MW;  B4E2FEB7EBA219C9 CRC64;


  Query Match             15.1%;  Score 142;  DB 1;  Length 192;
  Best Local Similarity   30.7%;  Pred. No. 0.0044;
  Matches   47;  Conservative   32;  Mismatches   44;  Indels   30;  Gaps    9;

Qy      34 WFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGLRNLYSW----SRVKGSLIIS 89
           |: |  ||:||||  :      : |: | :| : ::| |     :|    |:  | |:
Db      44 WWGITENAISDLGR-IGLPYNWVMNISLILGSVCLIIYG-----AWRFKKSKDIGWLLYM 97

Qy      90 MG-VFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISILDKSWI----------AVLL 138
           :| ||| |||:| |    :|: ||  ||:|:   |:|| :|  |::          :|
Db      98 LGSVFLGLIGIFPEGTN-LHYEVSWGFFVSM---FLAILLLSISFLIRGNKLGIVGLMLF 153

Qy     139 IIG-HIAMWYLHFASEIPRGAAIPELLAVFSFL 170
           ::|  :|:| |    :   | |: | :::  ||
Db     154 LLGVPLALWSL----KKFEGVAVAETISIVVFL 182


RESULT   4
O59239
ID   O59239      PRELIMINARY;      PRT;   183 AA.
AC   O59239;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 20.6 KDA PROTEIN PH1610.
GN   PH1610.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000006; BAA30722.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   183 AA;  20605 MW;  C0790C9DA464F92F CRC64;


  Query Match             13.7%;  Score 129.5;  DB 1;  Length 183;
  Best Local Similarity   29.4%;  Pred. No. 0.031;
  Matches   40;  Conservative   22;  Mismatches   47;  Indels   27;  Gaps    4;

Qy      28 AISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILI------VIVGLRNLYSWSR 81
           :|| : ||: : |||||||   |:|   |:| || :    :      :|:  :|
Db      30 SISNNPWFSFFKNALSDLGDPAKASHPWIYNYGLIVTSPFVLAFSIYLIIAAKN--KIQT 87

Qy      82 VKGSLIISMGVFLNLIGVFD---------EVYGWIHFLVSVLFF----------LSIIAY 122
           | |: |    :|| |||:|            | :| | : :| :           |||  :
Db      88 VGGAYISISSLFLALIGIFHAGTRPHVFVSTYFFIQFFLGMLIYGVGSKTAIRLTSIIIF 147

Qy     123 FIAISILDKSWIAVLL 138
           :|:  |   | :| |
Db     148 ILALLGLFIPWPSVAL 163


RESULT   5
Q9V1S7
ID   Q9V1S7      PRELIMINARY;      PRT;   156 AA.
AC   Q9V1S7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```

```
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 17.2 KDA PROTEIN.
GN   PAB0225.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ248284; CAB49272.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   156 AA;  17169 MW;  080E78DFDEC9DE8D CRC64;


  Query Match             13.4%;  Score 126.5;  DB 1;  Length 156;
  Best Local Similarity   25.3%;  Pred. No. 0.043;
  Matches   42;  Conservative   33;  Mismatches   64;  Indels   27;  Gaps    5;

Qy      17 SFSFPLTM-----IALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIV 71
           : |  ||:     :|| :  :|||:   |||||:| ::::    :||       : |: :
Db       4 ALSIALTLYFILGLALVVHENSWFSFSKNALSDMG-SLRNPKGWMFN-------VFIIGL 55

Qy      72 GLRNLYSWSRVKGSLIISMGVFLNLIGVFDE---VYGWIHFLVSVLFFLSIIAYFIAISI 128
           ||  | :   :|  ::     || |:||| |   ::     |  :| |  :: |
Db      56 GLLGLVTSLMLKRKILTLSMAFLVLVGVFPEEKPLHTPSAILTYILAFTDMVIYG----- 110

Qy     129 LDKSWIAVLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPFYI 174
                :  ::|      : |   :  | |||||:   | | : :
Db     111 ------GIWRVVGIGTFMVMLFLINLKVGLAIPELIGAASILAYIL 150


RESULT   6
Q9Y8U9
ID   Q9Y8U9      PRELIMINARY;      PRT;   165 AA.
AC   Q9Y8U9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 18.1 KDA PROTEIN APE2535.
GN   APE2535.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K., Kubota K., Nakamura Y.,
RA   Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000064; BAA81551.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   165 AA;  18081 MW;  82E9EC27C6669C15 CRC64;


  Query Match             12.5%;  Score 118;  DB 1;  Length 165;
  Best Local Similarity   30.4%;  Pred. No. 0.18;
  Matches   41;  Conservative   20;  Mismatches   52;  Indels   22;  Gaps    5;

Qy      32 SSWFNIWNNALSDLGHAVKSSVAPIFNLGLAIGGILIVIVGL---RNLYSWSRVKGS-LI 87
           | ||    :| ||||  ||: :  |:| || : |:|: :  |   |         || |:
Db       3 SDWFVFTEHAFSDLG-GVKARMPWIYNYGLILVGLLVTLYSLCPYRAAVERLEAFGSGLL 61

Qy      88 ISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISI----------------LDK 131
           :: |:|| ||||:       |  ||  ||| :    :|:|:                |
Db      62 LTAGLFLALIGVY-PAGTRPHVFVSTWFFLQMDMALVALSLGAYRACRNLANMAALALSL 120

Qy     132 SWIAVLLIIGHIAMW 146
           :   | ::||    |
Db     121 AAFPVYILIGITLGW 135


RESULT   7
Q9RQP4
ID   Q9RQP4      PRELIMINARY;      PRT;   512 AA.
AC   Q9RQP4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CYTOCHROME C OXIDASE SUBUNIT 1.
OS   Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC   Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC   Burkholderia.
OX   NCBI_TaxID=28450;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D2175;
RA   Sailan A.T., Zulkeflie Z., Mohamed M.R., Ariffin S.H., Embi M.N.;
RT   "The identification of cytochrome-c oxidase subunit 1 in Burkholderia
RT   pseudomallei.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF087002; AAF13732.1; -.
DR   INTERPRO; IPR000883; -.
DR   PFAM; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
SQ   SEQUENCE   512 AA;  56155 MW;  6B5433E207AD3EE0 CRC64;


  Query Match             11.6%;  Score 109.5;  DB 2;  Length 512;
  Best Local Similarity   22.9%;  Pred. No. 1.9;
  Matches   48;  Conservative   42;  Mismatches   61;  Indels   59;  Gaps   13;

Qy      15 ILSFSFPLTMIA---------LAISMS------SWFNIWNNALSDLGHAVKSSVAPIFNL 59
           :: : || |:||         ||: |       :|   |   |  ||  |  :: |: :|
Db      27 VMGYVFP-TLIAMGFGYAISELALKMPLVGRRWAWTGFW---LIGLGSVV--AMTPV-SL 79

Qy      60 GLA----------IG------GILIVIVG--------LRNLYSWSRVKGSLIISMGVFLN 95
           ||:          ||      |:::|:||          |||:| :      | : :| |
Db      80 GLSSVLYTFYPPMIGSPFYYLGVVLVVVGSWIWVALMSVNLYAWKKRNPGTPIPLAMFAN 139

Qy      96 LIGVFDEVYGW--IHFLVSVLFFLSIIAYFIAISI------LDKSWIAVLLIIGHIAMWY 147
           : | :  ::||  :   : :|| :  :|  :  :|      :  ||    ::   :   |
Db     140 VAGAY--LWGWTAVGAAIEILFQILPVAVGLKTTIDAGLARVFFSWTLHAIVYFWLIPAY 197

Qy     148 LHFASEIPR---GAAIPELLAVFSFLPFYI 174
           : : : :||   |    : :|  ||: | :
Db     198 IAYYTLVPRAIGGKLYSDGMARISFILFLV 227


RESULT   8
Q9KS98
ID   Q9KS98      PRELIMINARY;      PRT;   401 AA.
AC   Q9KS98;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   AMINO ACID ABC TRANSPORTER, PERMEASE PROTEIN.
GN   VC1361.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
```

```
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004215; AAF94519.1; -.
DR   TIGR; VC1361; -.
SQ   SEQUENCE   401 AA;  44207 MW;  76923D203A8C113A CRC64;


  Query Match            11.2%;  Score 105.5;  DB 2;  Length 401;
  Best Local Similarity   20.1%;  Pred. No. 2.8;
  Matches   46;  Conservative   38;  Mismatches   84;  Indels   61;  Gaps    7;

Qy       6 HSWMDMIIFILSF-SFPLTMIALAISMSSWFNIWNNALSDL----------------GHA 48
           ||    :::  :| |    ::|: :     : | |||::|                 |
Db      15 HSKHTNLLYNPTFRSVVFQLLAVCVLAFFLYTIVNNALTNLESRGIATGFAFLEQTAGFG 74

Qy      49 VKSSVAP-------------------IFNLGLAIGGILIVIVGLRNLYSWSRVKGSLII 88
           :  |: |                   :  ||: :  ||  ::|:  | |   :     :
Db      75 ISQSLIPYDETYSYGRTFLVGLLNTALVSVLGILLATILGFLIGIARLSSNWLISRLAAV 134

Qy      89 SMGVFLNLIGVFDEVYGWIHFLVSVL------------FFLSIIAYFIAISILDKS--WI 134
           :  | ||  :  ::: |   ::  |          |:|:|   ::|  | :    |:
Db     135 YIETFRNL-PLLLQIFFWYFVVLQALPSARESMHLGEWFYLNIKGLYLAKPIFESGSIWV 193

Qy     135 AVLLIIGHIAMWYLHF---------ASEIPRGAAIPELLAVFSFLPFYI 174
           | || | || | |           : | |  |  |  |  | :::
Db     194 LVALIAGIIACWVLSIWATNRQRLTGQQTPMGRYILLLCVTFPVLVYFL 242


RESULT   9
Q9RXQ6
ID   Q9RXQ6      PRELIMINARY;      PRT;   373 AA.
AC   Q9RXQ6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   DR0252.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001886; AAF09838.1; -.
DR   TIGR; DR0252; -.
DR   INTERPRO; IPR002549; -.
DR   PFAM; PF01594; UPF0118; 1.
SQ   SEQUENCE   373 AA;  40283 MW;  502D3FE29051E6F7 CRC64;


  Query Match            11.1%;  Score 104.5;  DB 2;  Length 373;
  Best Local Similarity   27.4%;  Pred. No. 3.1;
  Matches   43;  Conservative   29;  Mismatches   44;  Indels   41;  Gaps    9;

Qy      30 SMSSWF-NIWNNALSDLGHAVKSSVAPIFNLGLAIGGI-----LIVIVGLRNLYSWSRVK 83
           ::  |  || || :  | :|: ||   :||  ::|||:     ||::: :  :  ::||
Db     142 ALRDWIANIGNNIVPILQNAL-SSTGTLFNRLISIGGVIGQVLLILLLSVYLMLDYARVN 200

Qy      84 GSLIISMG--------VFLNLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISILDKSWIA 135
           ||: :          | :|||    | |::   :        :|| || | :    |:
Db     201 ASLLRAFPRPWQPKVLEFSDLIGT--AVGGYVRGQL-------LIAAFIGIFV----WLG 247

Qy     136 VLLIIGHIAMWYLHFASEIPRGAAIPELLAVFSFLPF 172
           | |:|            ||  |||  |   |: :|:
Db     248 -LTIVG-----------IPSAAAIGFLAGAFNIVPY 271


RESULT  10
Q9KNN1
ID   Q9KNN1      PRELIMINARY;      PRT;   600 AA.
AC   Q9KNN1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   THIOL:DISULFIDE INTERCHANGE PROTEIN DSBD.
GN   VC2701.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004336; AAF95841.1; -.
DR   TIGR; VC2701; -.
SQ   SEQUENCE   600 AA;  65209 MW;  23D01D76F730249C CRC64;


  Query Match            11.0%;  Score 103.5;  DB 2;  Length 600;
  Best Local Similarity   22.5%;  Pred. No. 5.5;
  Matches   42;  Conservative   40;  Mismatches   82;  Indels   23;  Gaps    7;

Qy       5 KHSWMDMIIFILSFSFPLTMIAL-AISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAI 63
           :| :: | : :|  :  |:|  | :: : |    | |:||:       |:  :| :|
Db     271 QHPYVLMGLSVLFVALALSMFGLYSLQLPSGVQTWLNSLSNAQQG--GSLPGVFAMGAIS 328

Qy      64 G----------GILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVF-----DEVYGW 106
           |          | |: :    :| : :    :| | ||: | |: ||      :   |
Db     329 GLVCSPCTTAPLSGALLYVAQSGDLLTGAVALYALAIGMGIPLILVAVFGNKLLPKAGNW 388

Qy     107 IHFLVSVLFFLSIIA-YFIAISILDKSWIAVL-LIIGHIAM-WYLHFASEIPRGAAIPEL 163
           :  : ::  |: : |  |:   |: : | :||   :|  |  |  |  : :| |     |
Db     389 MERVKTLFGFVLLAAPIFLLERIVPEFWSSVLWSALGLAAFGWLYHVKNSLPFGGWKQSL 448

Qy     164 LAVFSFL 170
           : : : |
Db     449 IGIVAIL 455


RESULT  11
Q9VGY0
ID   Q9VGY0      PRELIMINARY;      PRT;   293 AA.
AC   Q9VGY0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG4073 PROTEIN.
```

```
GN   CG4073.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003688; AAF54542.1; -.
DR   FLYBASE; FBgn0037827; CG4073.
SQ   SEQUENCE   293 AA;  32710 MW;  7B099E726C6916D0 CRC64;
```

```
  Query Match            10.9%;  Score 102.5;  DB 5;  Length 293;
  Best Local Similarity   24.5%;  Pred. No. 3.5;
  Matches   49;  Conservative   32;  Mismatches   62;  Indels   57;  Gaps    8;

Qy       5 KH-SWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLGLAI 63
           ||  |: :|  :|| |  : :: | :              | |  :  ||     | ||:
Db      73 KHLGWVYLIHAVLSASAVIQLVVLRL----------CNKDFGELISPSVPSFVWLLLAV 121

Qy      64 GGILI---------------------VIVGLRNLYSWSRVKG-------SLIISMGVFL 94
           | :||                     |||   | : |:|:         ||:::: | |
Db     122 GCVLIMAYVYLANQCPCNGLLAIVIVEVIVIFVNCHRWARLSMLWMTGVLSLVLALNVML 181

Qy      95 NLIGVFDEVYGWIHFLVSVLFFLSIIAYFIAISILDKSWIAVLLIIGHIAMWYLHFASEI 154
            |:|    ||  :  |   :|   |:  |  |:|:    ::: ||:      ||:    |
Db     182 YLMG----VYLPLKILPGSIFM--IVLTFCCIAIV----VSIYLIV------YLNGNRYI 225

Qy     155 PRGAAIPELLAVFSFLPFYI 174
           |  ::  |: | | : | |
Db     226 MRYVSMVSLIYVASLILFTI 245
```

```
RESULT  12
Q9XMT1
ID   Q9XMT1      PRELIMINARY;      PRT;   365 AA.
AC   Q9XMT1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ORF365.
OS   Tetrahymena pyriformis.
OG   Mitochondrion.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
OX   NCBI_TaxID=5908;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Edqvist J., Burger G., Gray M.W.;
RT   "Expression of mitochondrial protein-coding genes in Tetrahymena
RT   pyriformis.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N.,
RA   Lang B.F., Gray M.W.;
RT   "Complete sequence, gene content and organization of the mitochondrial
RT   genome of Tetrahymena pyriformis. Comparison with Paramecium aurelia
RT   mitochondrial DNA.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF160864; AAD41932.1; -.
DR   INTERPRO; IPR001064; -.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW   Mitochondrion.
SQ   SEQUENCE   365 AA;  43578 MW;  04FF18A0D1252D74 CRC64;
```

```
  Query Match            10.8%;  Score 101.5;  DB 8;  Length 365;
  Best Local Similarity   24.3%;  Pred. No. 4.9;
  Matches   37;  Conservative   27;  Mismatches   43;  Indels   45;  Gaps    7;

Qy      11 MIIFILSFSFPLT------MIALAISMSSWFNIWNNALSDL--------GHAVKSSVAPI 56
             ||:| : : | ||      ::   || ::  |  || :: |        |  :|  :|||
Db     211 MIVFSIFYIFTLTGTTEWSVVNFIISSNNQINYLNNKIAILFIAIILLIGFIIKLGIAPI 270

Qy      57 FNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFF 116
                              ||     ||        | : :|   |   :|||  |||
Db     271 ------------------QLYKIEIYKG---------LPFLSIF--FYTTFYFLVFFLFF 301

Qy     117 LSIIAYFIAISILDKSWIAVLLIIGHIAMWYL 148
           |  |::: ::::  |: :||::  | : |:
Db     302 SLIFIYYLS-ALINYYWL-LLLLVSIIGILYI 331
```

```
RESULT  13
Q9KLZ9
ID   Q9KLZ9      PRELIMINARY;      PRT;   360 AA.
AC   Q9KLZ9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PEPTIDE ABC TRANSPORTER, PERMEASE PROTEIN, PUTATIVE.
GN   VCA0590.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
```

```
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004389; AAF96492.1; -.
DR   TIGR; VCA0590; -.
SQ   SEQUENCE   360 AA;  40400 MW;  FC19A38223D26E2A CRC64;


  Query Match           10.5%;  Score 98.5;  DB 2;  Length 360;
  Best Local Similarity  25.8%;  Pred. No. 7.8;
  Matches   34;  Conservative   29;  Mismatches   52;  Indels   17;  Gaps    6;

Qy      1 MSLNKHSWMDMIIFILSFSFPLTMIALAISMSSWFNIWNNALSDLGHAVKSSVAPIFNLG 60
          :|::   |  ::|:::  | ||  |: ||   | |::|::|:  :|:|:          |
Db    131 VSISLGLWSTLLIYLI--SIPLG-ISKAIHHGSRFDVWSSAVVIIGYAIP---------G 178

Qy     61 LAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYGWIHFLVSVLFFLSII 120
           |||::    | :||  ::| :  :         | |  | | |  :  |    :|
Db    179 FLFAIILIILFASGNYFSWFPLRGLVSDNFASLPWYQQVLD--YFW-HLTLPTL--AMVI 233

Qy    121 AYFIAISILDKS 132
           |  :|:| |:
Db    234 GGFATLSMLTKN 245


RESULT  14
Q9ZZU5
ID   Q9ZZU5      PRELIMINARY;      PRT;   231 AA.
AC   Q9ZZU5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ATP SYNTHASE A CHAIN (EC 3.6.1.34).
GN   ATPASE 6.
OS   Littorina saxatilis.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC   Neotaenioglossa; Littorinoidea; Littorinidae; Littorina.
OX   NCBI_TaxID=31220;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilding C.S., Mill P.J., Grahame J.;
RT   "Partial sequence of the mitochondrial genome of Littorina saxatilis:
RT   relevance to gastropod phylogenetics.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AJ132137; CAA10596.1; -.
DR   INTERPRO; IPR000568; -.
DR   PFAM; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   Mitochondrion; Hydrogen ion transport; CF(0); Transmembrane.
SQ   SEQUENCE   231 AA;  25478 MW;  7EEAF487BC53A67A CRC64;


  Query Match           10.4%;  Score 97.5;  DB 8;  Length 231;
  Best Local Similarity  24.2%;  Pred. No. 6.2;
  Matches   53;  Conservative   34;  Mismatches   63;  Indels   69;  Gaps   12;

Qy      2 SLNKHSWMDMIIFILSFSFPLTMIALAISMSSWFN--IWNNALSDLGHAVKSSVAPIFNL 59
          | : :: : | :::| ::| | :| |  | | |     |   :|     : | |   |
Db      8 SFDDNNQVFMSLYMLMWAFSLVII-LIFSSSYWVTGPRWAPIISLFKDTISSQVFRSF-- 64

Qy     60 GLAIGGILIVIVGLRNLYSWSRVKGSLIISMGVFLNLIGVFDEVYG---------- 105
             ||::|| | || ||                :|||| |:   |:
Db     65 GLSLGGFLNVITGL--------------FLFLIFLNLAGLIPYVFSPTSHLAVSLSLGLP 110

Qy    106 -WIHFLVSVLFF--LSIIAYFI---------------AISILDKS-WIAVLLI----IG 141
           |:  |:|  ||  |::|  :                ::||| :   ::| |:     |
Db    111 LWLSLLISGTFFNPSSVVAGLLPMGAPAPLNPFLVLIESVSILVRPITLSVRLMANMSAG 170

Qy    142 HIAMW----YLHFASEIPRGA-------AIPELLAVFSF 169
           || :     |:  :  :| ||       :|  |  :| |
Db    171 HIVLTLIGNYMTTSLFMP-GAFSMLLLISIQVLYTIFEF 208


RESULT  15
Q9KM39
ID   Q9KM39      PRELIMINARY;      PRT;   360 AA.
AC   Q9KM39;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL PROTEIN VCA0550.
GN   VCA0550.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004386; AAF96452.1; -.
DR   TIGR; VCA0550; -.
KW   Hypothetical protein.
SQ   SEQUENCE   360 AA;  40265 MW;  999D1FFB2FD08887 CRC64;


  Query Match           10.4%;  Score 97.5;  DB 2;  Length 360;
  Best Local Similarity  25.9%;  Pred. No. 9.2;
  Matches   45;  Conservative   29;  Mismatches   49;  Indels   51;  Gaps   10;

Qy      8 WMDMIIFILS--FSFPLTMIALAI----SMSSW-----FNIWNNALS-------DLGHAV 49
          | |::| |:   | | |:::  : :          |     || : |  :       | || :
Db    197 WQDVVISIMGPCFGFLLSLLLVGVYWVTDSPFWAALAVFNAFLNLFNLLPVLPLDGGHIL 256

Qy     50 KSSVAPIFNLGLAIGGILIVIVGLRNLYSWSRVKGSLIISMGV----FLNLIGVFDEVYG 105
          ||     |::   :| || |:  |  :         :|  |:|:    || |:|  | |:
Db    257 KS---VSFSMNSKMGVILCVLAILGGI--------ALSYSLGLTLFGFLLLMGALDIVFE 305

Qy    106 W-----IHFL--------VSVLFFLSIIAYFIAI-----SILDKSWIAVLLIIG 141
          |      | |        || ::: ::::  |||     |  |      |||:|
Db    306 WRQRHHSHLLPLNRYAQMVSTIWYFALVSGLIAIIIGFASTGDTLLSLPLLILG 359


Search completed: May 30, 2001, 15:25:50
Job time: 120 sec
```